Our Ref.: **CL001208**

U.S. PATENT APPLICATION

Inventor(s): Hongyu ZHANG
Ming-Hui WEI
Karen A. KETCHUM
Valentina DI FRANCESCO
Ellen M. BEASLEY

Invention: ISOLATED HUMAN TRANSPORTER PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS, AND USES THEREOF

CELERA GENOMICS CORPORATION.
45 WEST GUDE DR., C2-4#20
ROCKVILLE, MD 20850
(240) 453-3067
Fax (240)-453-3084

SPECIFICATION

ISOLATED HUMAN TRANSPORTER PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS, AND USES THEREOF

FIELD OF THE INVENTION

The present invention is in the field of transporter proteins that are related to the ion channel transporter subfamily, recombinant DNA molecules, and protein production. The present invention specifically provides novel peptides and proteins that effect ligand transport and nucleic acid molecules encoding such peptide and protein molecules, all of which are useful in the development of human therapeutics and diagnostic compositions and methods.

BACKGROUND OF THE INVENTION

Transporters

Transporter proteins regulate many different functions of a cell, including cell proliferation, differentiation, and signaling processes, by regulating the flow of molecules such as ions and macromolecules, into and out of cells. Transporters are found in the plasma membranes of virtually every cell in eukaryotic organisms. Transporters mediate a variety of cellular functions including regulation of membrane potentials and absorption and secretion of molecules and ion across cell membranes. When present in intracellular membranes of the Golgi apparatus and endocytic vesicles, transporters, such as chloride channels, also regulate organelle pH. For a review, see Greger, R. (1988) Annu. Rev. Physiol. 50:111-122.

Transporters are generally classified by structure and the type of mode of action. In addition, transporters are sometimes classified by the molecule type that is transported, for example, sugar transporters, chlorine channels, potassium channels, etc. There may be many classes of channels for transporting a single type of molecule (a detailed review of channel types can be found at Alexander, S.P.H. and J.A. Peters: Receptor and

transporter nomenclature supplement. Trends Pharmacol. Sci., Elsevier, pp. 65-68 (1997) and http://www-biology.ucsd.edu/~msaier/transport/titlepage2.html.

The following general classification scheme is known in the art and is followed in the present discoveries.

Channel-type transporters. Transmembrane channel proteins of this class are ubiquitously found in the membranes of all types of organisms from bacteria to higher eukaryotes. Transport systems of this type catalyze facilitated diffusion (by an energy-independent process) by passage through a transmembrane aqueous pore or channel without evidence for a carrier-mediated mechanism. These channel proteins usually consist largely of a-helical spanners, although b-strands may also be present and may even comprise the channel. However, outer membrane porin-type channel proteins are excluded from this class and are instead included in class 9.

Carrier-type transporters. Transport systems are included in this class if they utilize a carrier-mediated process to catalyze uniport (a single species is transported by facilitated diffusion), antiport (two or more species are transported in opposite directions in a tightly coupled process, not coupled to a direct form of energy other than chemiosmotic energy) and/or symport (two or more species are transported together in the same direction in a tightly coupled process, not coupled to a direct form of energy other than chemiosmotic energy).

Pyrophosphate bond hydrolysis-driven active transporters. Transport systems are included in this class if they hydrolyze pyrophosphate or the terminal pyrophosphate bond in ATP or another nucleoside triphosphate to drive the active uptake and/or extrusion of a solute or solutes. The transport protein may or may not be transiently phosphorylated, but the substrate is not phosphorylated.

PEP-dependent, phosphoryl transfer-driven group translocators. Transport systems of the bacterial phosphoenolpyruvate:sugar phosphotransferase system are included in this class. The product of the reaction, derived from extracellular sugar, is a cytoplasmic sugar-phosphate.

Decarboxylation-driven active transporters. Transport systems that drive solute (e.g., ion) uptake or extrusion by decarboxylation of a cytoplasmic substrate are included in this class.

Oxidoreduction-driven active transporters. Transport systems that drive transport of a solute (e.g., an ion) energized by the flow of electrons from a reduced substrate to an oxidized substrate are included in this class.

Light-driven active transporters. Transport systems that utilize light energy to drive transport of a solute (e.g., an ion) are included in this class.

Mechanically-driven active transporters. Transport systems are included in this class if they drive movement of a cell or organelle by allowing the flow of ions (or other solutes) through the membrane down their electrochemical gradients.

Outer-membrane porins (of b-structure). These proteins form transmembrane pores or channels that usually allow the energy independent passage of solutes across a membrane. The transmembrane portions of these proteins consist exclusively of b-strands that form a b-barrel. These porin-type proteins are found in the outer membranes of Gram-negative bacteria, mitochondria and eukaryotic plastids.

Methyltransferase-driven active transporters. A single characterized protein currently falls into this category, the Na+-transporting methyltetrahydromethanopterin:coenzyme M methyltransferase.

Non-ribosome-synthesized channel-forming peptides or peptide-like molecules. These molecules, usually chains of L- and D-amino acids as well as other small molecular building blocks such as lactate, form oligomeric transmembrane ion channels. Voltage may induce channel formation by promoting assembly of the transmembrane channel. These peptides are often made by bacteria and fungi as agents of biological warfare.

Non-Proteinaceous Transport Complexes. Ion conducting substances in biological membranes that do not consist of or are not derived from proteins or peptides fall into this category.

Functionally characterized transporters for which sequence data are lacking. Transporters of particular physiological significance will be included in this category even though a family assignment cannot be made.

Putative transporters in which no family member is an established transporter. Putative transport protein families are grouped under this number and will either be classified elsewhere when the transport function of a member becomes established, or

will be eliminated from the TC classification system if the proposed transport function is disproven. These families include a member or members for which a transport function has been suggested, but evidence for such a function is not yet compelling.

Auxiliary transport proteins. Proteins that in some way facilitate transport across one or more biological membranes but do not themselves participate directly in transport are included in this class. These proteins always function in conjunction with one or more transport proteins. They may provide a function connected with energy coupling to transport, play a structural role in complex formation or serve a regulatory function.

Transporters of unknown classification. Transport protein families of unknown classification are grouped under this number and will be classified elsewhere when the transport process and energy coupling mechanism are characterized. These families include at least one member for which a transport function has been established, but either the mode of transport or the energy coupling mechanism is not known.

Ion channels

An important type of transporter is the ion channel. Ion channels regulate many different cell proliferation, differentiation, and signaling processes by regulating the flow of ions into and out of cells. Ion channels are found in the plasma membranes of virtually every cell in eukaryotic organisms. Ion channels mediate a variety of cellular functions including regulation of membrane potentials and absorption and secretion of ion across epithelial membranes. When present in intracellular membranes of the Golgi apparatus and endocytic vesicles, ion channels, such as chloride channels, also regulate organelle pH. For a review, see Greger, R. (1988) Annu. Rev. Physiol. 50:111-122.

Ion channels are generally classified by structure and the type of mode of action. For example, extracellular ligand gated channels (ELGs) are comprised of five polypeptide subunits, with each subunit having 4 membrane spanning domains, and are activated by the binding of an extracellular ligand to the channel. In addition, channels are sometimes classified by the ion type that is transported, for example, chlorine channels, potassium channels, etc. There may be many classes of channels for transporting a single type of ion (a detailed review of channel types can be found at Alexander, S.P.H. and J.A. Peters (1997). Receptor and ion channel nomenclature

supplement. Trends Pharmacol. Sci., Elsevier, pp. 65-68 and http://www-biology.ucsd.edu/~msaier/transport/toc.html.

There are many types of ion channels based on structure. For example, many ion channels fall within one of the following groups: extracellular ligand-gated channels (ELG), intracellular ligand-gated channels (ILG), inward rectifying channels (INR), intercellular (gap junction) channels, and voltage gated channels (VIC). There are additionally recognized other channel families based on ion-type transported, cellular location and drug sensitivity. Detailed information on each of these, their activity, ligand type, ion type, disease association, drugability, and other information pertinent to the present invention, is well known in the art.

Extracellular ligand-gated channels, ELGs, are generally comprised of five polypeptide subunits, Unwin, N. (1993), Cell 72: 31-41; Unwin, N. (1995), Nature 373: 37-43; Hucho, F., et al., (1996) J. Neurochem. 66: 1781-1792; Hucho, F., et al., (1996) Eur. J. Biochem. 239: 539-557; Alexander, S.P.H. and J.A. Peters (1997), Trends Pharmacol. Sci., Elsevier, pp. 4-6; 36-40; 42-44; and Xue, H. (1998) J. Mol. Evol. 47: 323-333. Each subunit has 4 membrane spanning regions: this serves as a means of identifying other members of the ELG family of proteins. ELG bind a ligand and in response modulate the flow of ions. Examples of ELG include most members of the neurotransmitter-receptor family of proteins, e.g., GABAI receptors. Other members of this family of ion channels include glycine receptors, ryandyne receptors, and ligand gated calcium channels.

Chloride Intracellular Channels (CLIC)

The novel human protein, and encoding gene, provided by the present invention is related to the ion channel family in general and the chloride intracellular channel (CLIC) in particular. Furthermore, the protein/cDNA of the present invention may be an alternative splice form of a CLIC2 protein/gene provided in Genbank gi4557020. CLIC2, along with CLIC1, localize intracellularly in the cytoplasm and nucleus, unlike other chloride channels. Direct associations have been found between numerous chloride channel genes and numerous hereditary diseases (Heiss *et al., Genomics* 45: 224-228, 1997); thus, CLIC2 and other novel human CLICs are valuable candidates for many

disorders. The CLIC2 gene maps to the candidate region of chromosome X for incontinentia pigmentia, which is a disorder characterized by abnormalities of tissues/organs that develop from ectoderm and neuroectoderm (Rogner *et al., Genome Res.* 6: 922-934, 1996).

Chloride channels are important in a wide variety of physiological processes in humans. For example, chloride channels regulate fundamental cellular processes such as cell membrane potential stabilization, transepithelial transport, intracellular pH maintenance, and cell volume mantenance.

Due to their importance in regulating fundamental cellular processes, novel human CLIC proteins/genes, such as provided by the present invention, are valuable as potential targets for the development of therapeutics to treat a wide variety of diseases/disorders such as cancer. Furthermore, SNPs in CLIC genes, such as provided by the present invention, may serve as valuable markers for the diagnosis, prognosis, prevention, and/or treatment of these diseases/disorders.

Using the information provided by the present invention, reagents such as probes/primers for detecting the SNPs or the expression of the protein/gene provided herein may be readily developed and, if desired, incorporated into kit formats such as nucleic acid arrays, primer extension reactions coupled with mass spec detection (for SNP detection), or TaqMan PCR assays (Applied Biosystems, Foster City, CA).

The Voltage-gated Ion Channel (VIC) Superfamily

Proteins of the VIC family are ion-selective channel proteins found in a wide range of bacteria, archaea and eukaryotes Hille, B. (1992), Chapter 9: Structure of channel proteins; Chapter 20: Evolution and diversity. In: Ionic Channels of Excitable Membranes, 2nd Ed., Sinaur Assoc. Inc., Pubs., Sunderland, Massachusetts; Sigworth, F.J. (1993), Quart. Rev. Biophys. 27: 1-40; Salkoff, L. and T. Jegla (1995), Neuron 15: 489-492; Alexander, S.P.H. et al., (1997), Trends Pharmacol. Sci., Elsevier, pp. 76-84; Jan, L.Y. et al., (1997), Annu. Rev. Neurosci. 20: 91-123; Doyle, D.A, et al., (1998) Science 280: 69-77; Terlau, H. and W. Stühmer (1998), Naturwissenschaften 85: 437-444. They are often homo- or heterooligomeric structures with several dissimilar subunits (e.g., a1-a2-d-b Ca^{2+} channels, ab_1b_2 Na^+ channels or $(a)_4$-b K^+ channels), but

the channel and the primary receptor is usually associated with the a (or a1) subunit. Functionally characterized members are specific for K^+, Na^+ or Ca^{2+}. The K^+ channels usually consist of homotetrameric structures with each a-subunit possessing six transmembrane spanners (TMSs). The a1 and a subunits of the Ca^{2+} and Na^+ channels, respectively, are about four times as large and possess 4 units, each with 6 TMSs separated by a hydrophilic loop, for a total of 24 TMSs. These large channel proteins form heterotetra-unit structures equivalent to the homotetrameric structures of most K^+ channels. All four units of the Ca^{2+} and Na^+ channels are homologous to the single unit in the homotetrameric K^+ channels. Ion flux via the eukaryotic channels is generally controlled by the transmembrane electrical potential (hence the designation, voltage-sensitive) although some are controlled by ligand or receptor binding.

Several putative K^+-selective channel proteins of the VIC family have been identified in prokaryotes. The structure of one of them, the KcsA K^+ channel of *Streptomyces lividans*, has been solved to 3.2 Å resolution. The protein possesses four identical subunits, each with two transmembrane helices, arranged in the shape of an inverted teepee or cone. The cone cradles the "selectivity filter" P domain in its outer end. The narrow selectivity filter is only 12 Å long, whereas the remainder of the channel is wider and lined with hydrophobic residues. A large water-filled cavity and helix dipoles stabilize K^+ in the pore. The selectivity filter has two bound K^+ ions about 7.5 Å apart from each other. Ion conduction is proposed to result from a balance of electrostatic attractive and repulsive forces.

In eukaryotes, each VIC family channel type has several subtypes based on pharmacological and electrophysiological data. Thus, there are five types of Ca^{2+} channels (L, N, P, Q and T). There are at least ten types of K^+ channels, each responding in different ways to different stimuli: voltage-sensitive [Ka, Kv, Kvr, Kvs and Ksr], Ca^{2+}-sensitive [BK_{Ca}, IK_{Ca} and SK_{Ca}] and receptor-coupled [K_M and K_{ACh}]. There are at least six types of Na^+ channels (I, II, III, μ1, H1 and PN3). Tetrameric channels from both prokaryotic and eukaryotic organisms are known in which each a-subunit possesses 2 TMSs rather than 6, and these two TMSs are homologous to TMSs 5 and 6 of the six TMS unit found in the voltage-sensitive channel proteins. KcsA of *S. lividans* is an example of such a 2 TMS channel protein. These channels may include the K_{Na} (Na^+-

activated) and K_{Vol} (cell volume-sensitive) K^+ channels, as well as distantly related channels such as the Tok1 K^+ channel of yeast, the TWIK-1 inward rectifier K^+ channel of the mouse and the TREK-1 K^+ channel of the mouse. Because of insufficient sequence similarity with proteins of the VIC family, inward rectifier K^+ IRK channels (ATP-regulated; G-protein-activated) which possess a P domain and two flanking TMSs are placed in a distinct family. However, substantial sequence similarity in the P region suggests that they are homologous. The b, g and d subunits of VIC family members, when present, frequently play regulatory roles in channel activation/deactivation.

The Epithelial Na^+ Channel (ENaC) Family

The ENaC family consists of over twenty-four sequenced proteins (Canessa, C.M., et al., (1994), Nature 367: 463-467, Le, T. and M.H. Saier, Jr. (1996), Mol. Membr. Biol. 13: 149-157; Garty, H. and L.G. Palmer (1997), Physiol. Rev. 77: 359-396; Waldmann, R., et al., (1997), Nature 386: 173-177; Darboux, I., et al., (1998), J. Biol. Chem. 273: 9424-9429; Firsov, D., et al., (1998), EMBO J. 17: 344-352; Horisberger, J.-D. (1998). Curr. Opin. Struc. Biol. 10: 443-449). All are from animals with no recognizable homologues in other eukaryotes or bacteria. The vertebrate ENaC proteins from epithelial cells cluster tightly together on the phylogenetic tree: voltage-insensitive ENaC homologues are also found in the brain. Eleven sequenced *C. elegans* proteins, including the degenerins, are distantly related to the vertebrate proteins as well as to each other. At least some of these proteins form part of a mechano-transducing complex for touch sensitivity. The homologous *Helix aspersa* (FMRF-amide)-activated Na^+ channel is the first peptide neurotransmitter-gated ionotropic receptor to be sequenced.

Protein members of this family all exhibit the same apparent topology, each with N- and C-termini on the inside of the cell, two amphipathic transmembrane spanning segments, and a large extracellular loop. The extracellular domains contain numerous highly conserved cysteine residues. They are proposed to serve a receptor function.

Mammalian ENaC is important for the maintenance of Na^+ balance and the regulation of blood pressure. Three homologous ENaC subunits, alpha, beta, and gamma, have been shown to assemble to form the highly Na^+-selective channel. The stoichiometry of the three subunits is $alpha_2$, beta1, gamma1 in a heterotetrameric architecture.

The Glutamate-gated Ion Channel (GIC) Family of Neurotransmitter Receptors

Members of the GIC family are heteropentameric complexes in which each of the 5 subunits is of 800-1000 amino acyl residues in length (Nakanishi, N., et al, (1990), Neuron 5: 569-581; Unwin, N. (1993), Cell 72: 31-41; Alexander, S.P.H. and J.A. Peters (1997) Trends Pharmacol. Sci., Elsevier, pp. 36-40). These subunits may span the membrane three or five times as putative a-helices with the N-termini (the glutamate-binding domains) localized extracellularly and the C-termini localized cytoplasmically. They may be distantly related to the ligand-gated ion channels, and if so, they may possess substantial b-structure in their transmembrane regions. However, homology between these two families cannot be established on the basis of sequence comparisons alone. The subunits fall into six subfamilies: a, b, g, d, e and z.

The GIC channels are divided into three types: (1) a-amino-3-hydroxy-5-methyl-4-isoxazole propionate (AMPA)-, (2) kainate- and (3) N-methyl-D-aspartate (NMDA)-selective glutamate receptors. Subunits of the AMPA and kainate classes exhibit 35-40% identity with each other while subunits of the NMDA receptors exhibit 22-24% identity with the former subunits. They possess large N-terminal, extracellular glutamate-binding domains that are homologous to the periplasmic glutamine and glutamate receptors of ABC-type uptake permeases of Gram-negative bacteria. All known members of the GIC family are from animals. The different channel (receptor) types exhibit distinct ion selectivities and conductance properties. The NMDA-selective large conductance channels are highly permeable to monovalent cations and Ca^{2+}. The AMPA- and kainate-selective ion channels are permeable primarily to monovalent cations with only low permeability to Ca^{2+}.

The Chloride Channel (ClC) Family

The ClC family is a large family consisting of dozens of sequenced proteins derived from Gram-negative and Gram-positive bacteria, cyanobacteria, archaea, yeast, plants and animals (Steinmeyer, K., et al., (1991), Nature 354: 301-304; Uchida, S., et al., (1993), J. Biol. Chem. 268: 3821-3824; Huang, M.-E., et al., (1994), J. Mol. Biol. 242: 595-598; Kawasaki, M., et al, (1994), Neuron 12: 597-604; Fisher, W.E., et al., (1995), Genomics. 29:598-606; and Foskett, J.K. (1998), Annu. Rev. Physiol. 60: 689-717).

These proteins are essentially ubiquitous, although they are not encoded within genomes of *Haemophilus influenzae*, *Mycoplasma genitalium,* and *Mycoplasma pneumoniae*. Sequenced proteins vary in size from 395 amino acyl residues *(M. jannaschii)* to 988 residues (man). Several organisms contain multiple ClC family paralogues. For example, *Synechocystis* has two paralogues, one of 451 residues in length and the other of 899 residues. *Arabidopsis thaliana* has at least four sequenced paralogues, (775-792 residues), humans also have at least five paralogues (820-988 residues), and *C. elegans* also has at least five (810-950 residues). There are nine known members in mammals, and mutations in three of the corresponding genes cause human diseases. *E. coli*, *Methanococcus jannaschii* and *Saccharomyces cerevisiae* only have one ClC family member each. With the exception of the larger *Synechocystis* paralogue, all bacterial proteins are small (395-492 residues) while all eukaryotic proteins are larger (687-988 residues). These proteins exhibit 10-12 putative transmembrane a-helical spanners (TMSs) and appear to be present in the membrane as homodimers. While one member of the family, *Torpedo* ClC-O, has been reported to have two channels, one per subunit, others are believed to have just one.

All functionally characterized members of the ClC family transport chloride, some in a voltage-regulated process. These channels serve a variety of physiological functions (cell volume regulation; membrane potential stabilization; signal transduction; transepithelial transport, etc.). Different homologues in humans exhibit differing anion selectivities, i.e., ClC4 and ClC5 share a $NO_3^- > Cl^- > Br^- > I^-$ conductance sequence, while ClC3 has an $I^- > Cl^-$ selectivity. The ClC4 and ClC5 channels and others exhibit outward rectifying currents with currents only at voltages more positive than +20mV.

Animal Inward Rectifier K^+ Channel (IRK-C) Family

IRK channels possess the "minimal channel-forming structure" with only a P domain, characteristic of the channel proteins of the VIC family, and two flanking transmembrane spanners (Shuck, M.E., et al., (1994), J. Biol. Chem. 269: 24261-24270; Ashen, M.D., et al., (1995), Am. J. Physiol. 268: H506-H511; Salkoff, L. and T. Jegla (1995), Neuron 15: 489-492; Aguilar-Bryan, L., et al., (1998), Physiol. Rev. 78: 227-245; Ruknudin, A., et al., (1998), J. Biol. Chem. 273: 14165-14171). They may exist in the membrane as homo- or heterooligomers. They have a greater tendency to let K^+ flow into

the cell than out. Voltage-dependence may be regulated by external K^+, by internal Mg^{2+}, by internal ATP and/or by G-proteins. The P domains of IRK channels exhibit limited sequence similarity to those of the VIC family, but this sequence similarity is insufficient to establish homology. Inward rectifiers play a role in setting cellular membrane potentials, and the closing of these channels upon depolarization permits the occurrence of long duration action potentials with a plateau phase. Inward rectifiers lack the intrinsic voltage sensing helices found in VIC family channels. In a few cases, those of Kir1.1a and Kir6.2, for example, direct interaction with a member of the ABC superfamily has been proposed to confer unique functional and regulatory properties to the heteromeric complex, including sensitivity to ATP. The SUR1 sulfonylurea receptor (spQ09428) is the ABC protein that regulates the Kir6.2 channel in response to ATP, and CFTR may regulate Kir1.1a. Mutations in SUR1 are the cause of familial persistent hyperinsulinemic hypoglycemia in infancy (PHHI), an autosomal recessive disorder characterized by unregulated insulin secretion in the pancreas.

ATP-gated Cation Channel (ACC) Family

Members of the ACC family (also called P2X receptors) respond to ATP, a functional neurotransmitter released by exocytosis from many types of neurons (North, R.A. (1996), Curr. Opin. Cell Biol. 8: 474-483; Soto, F., M. Garcia-Guzman and W. Stühmer (1997), J. Membr. Biol. 160: 91-100). They have been placed into seven groups ($P2X_1$ - $P2X_7$) based on their pharmacological properties. These channels, which function at neuron-neuron and neuron-smooth muscle junctions, may play roles in the control of blood pressure and pain sensation. They may also function in lymphocyte and platelet physiology. They are found only in animals.

The proteins of the ACC family are quite similar in sequence (>35% identity), but they possess 380-1000 amino acyl residues per subunit with variability in length localized primarily to the C-terminal domains. They possess two transmembrane spanners, one about 30-50 residues from their N-termini, the other near residues 320-340. The extracellular receptor domains between these two spanners (of about 270 residues) are well conserved with numerous conserved glycyl and cysteyl residues. The hydrophilic C-termini vary in length from 25 to 240 residues. They resemble the topologically similar epithelial Na^+ channel (ENaC) proteins in possessing (a) N- and C-termini localized

intracellularly, (b) two putative transmembrane spanners, (c) a large extracellular loop domain, and (d) many conserved extracellular cysteyl residues. ACC family members are, however, not demonstrably homologous with them. ACC channels are probably hetero- or homomultimers and transport small monovalent cations (Me^+). Some also transport Ca^{2+}; a few also transport small metabolites.

The Ryanodine-Inositol 1,4,5-triphosphate Receptor Ca^{2+} Channel (RIR-CaC) Family

Ryanodine (Ry)-sensitive and inositol 1,4,5-triphosphate (IP3)-sensitive Ca^{2+}-release channels function in the release of Ca^{2+} from intracellular storage sites in animal cells and thereby regulate various Ca^{2+} -dependent physiological processes (Hasan, G. et al., (1992) Development 116: 967-975; Michikawa, T., et al., (1994), J. Biol. Chem. 269: 9184-9189; Tunwell, R.E.A., (1996), Biochem. J. 318: 477-487; Lee, A.G. (1996) *Biomembranes,* Vol. 6, Transmembrane Receptors and Channels (A.G. Lee, ed.), JAI Press, Denver, CO., pp 291-326; Mikoshiba, K., et al., (1996) J. Biochem. Biomem. 6: 273-289). Ry receptors occur primarily in muscle cell sarcoplasmic reticular (SR) membranes, and IP3 receptors occur primarily in brain cell endoplasmic reticular (ER) membranes where they effect release of Ca^{2+} into the cytoplasm upon activation (opening) of the channel.

The Ry receptors are activated as a result of the activity of dihydropyridine-sensitive Ca^{2+} channels. The latter are members of the voltage-sensitive ion channel (VIC) family. Dihydropyridine-sensitive channels are present in the T-tubular systems of muscle tissues.

Ry receptors are homotetrameric complexes with each subunit exhibiting a molecular size of over 500,000 daltons (about 5,000 amino acyl residues). They possess C-terminal domains with six putative transmembrane a -helical spanners (TMSs). Putative pore-forming sequences occur between the fifth and sixth TMSs as suggested for members of the VIC family. The large N-terminal hydrophilic domains and the small C-terminal hydrophilic domains are localized to the cytoplasm. Low resolution 3-dimensional structural data are available. Mammals possess at least three isoforms that probably arose by gene duplication and divergence before divergence of the mammalian species. Homologues are present in humans and *Caenorabditis elegans.*

IP_3 receptors resemble Ry receptors in many respects. (1) They are homotetrameric complexes with each subunit exhibiting a molecular size of over 300,000 daltons (about 2,700 amino acyl residues). (2) They possess C-terminal channel domains that are homologous to those of the Ry receptors. (3) The channel domains possess six putative TMSs and a putative channel lining region between TMSs 5 and 6. (4) Both the large N-terminal domains and the smaller C-terminal tails face the cytoplasm. (5) They possess covalently linked carbohydrate on extracytoplasmic loops of the channel domains. (6) They have three currently recognized isoforms (types 1, 2, and 3) in mammals which are subject to differential regulation and have different tissue distributions.

IP_3 receptors possess three domains: N-terminal IP_3-binding domains, central coupling or regulatory domains and C-terminal channel domains. Channels are activated by IP_3 binding, and like the Ry receptors, the activities of the IP_3 receptor channels are regulated by phosphorylation of the regulatory domains, catalyzed by various protein kinases. They predominate in the endoplasmic reticular membranes of various cell types in the brain but have also been found in the plasma membranes of some nerve cells derived from a variety of tissues.

The channel domains of the Ry and IP_3 receptors comprise a coherent family that in spite of apparent structural similarities, do not show appreciable sequence similarity of the proteins of the VIC family. The Ry receptors and the IP_3 receptors cluster separately on the RIR-CaC family tree. They both have homologues in *Drosophila.* Based on the phylogenetic tree for the family, the family probably evolved in the following sequence: (1) A gene duplication event occurred that gave rise to Ry and IP_3 receptors in invertebrates. (2) Vertebrates evolved from invertebrates. (3) The three isoforms of each receptor arose as a result of two distinct gene duplication events. (4) These isoforms were transmitted to mammals before divergence of the mammalian species.

The Organellar Chloride Channel (O-ClC) Family

Proteins of the O-ClC family are voltage-sensitive chloride channels found in intracellular membranes but not the plasma membranes of animal cells (Landry, D, et al., (1993), J. Biol. Chem. 268: 14948-14955; Valenzuela, Set al., (1997), J. Biol. Chem. 272: 12575-12582; and Duncan, R.R., et al., (1997), J. Biol. Chem. 272: 23880-23886).

They are found in human nuclear membranes, and the bovine protein targets to the microsomes, but not the plasma membrane, when expressed in *Xenopus laevis* oocytes. These proteins are thought to function in the regulation of the membrane potential and in transepithelial ion absorption and secretion in the kidney. They possess two putative transmembrane a-helical spanners (TMSs) with cytoplasmic N- and C-termini and a large luminal loop that may be glycosylated. The bovine protein is 437 amino acyl residues in length and has the two putative TMSs at positions 223-239 and 367-385. The human nuclear protein is much smaller (241 residues). A *C. elegans* homologue is 260 residues long.

Transporter proteins, particularly members of the ion channel subfamily, are a major target for drug action and development. Accordingly, it is valuable to the field of pharmaceutical development to identify and characterize previously unknown transport proteins. The present invention advances the state of the art by providing previously unidentified human transport proteins.

SUMMARY OF THE INVENTION

The present invention is based in part on the identification of amino acid sequences of human transporter peptides and proteins that are related to the ion channel transporter subfamily, as well as allelic variants and other mammalian orthologs thereof. These unique peptide sequences, and nucleic acid sequences that encode these peptides, can be used as models for the development of human therapeutic targets, aid in the identification of therapeutic proteins, and serve as targets for the development of human therapeutic agents that modulate transporter activity in cells and tissues that express the transporter. Experimental data as provided in Figure 1 indicates expression in humans in the uterus, lung, germ cells, liver, parathyroid gland, prostate, and placenta.

DESCRIPTION OF THE FIGURE SHEETS

FIGURE 1 provides the nucleotide sequence of a cDNA molecule that encodes the transporter protein of the present invention. (SEQ ID NO:1) In addition structure and functional information is provided, such as ATG start, stop and tissue distribution, where

available, that allows one to readily determine specific uses of inventions based on this molecular sequence. Experimental data as provided in Figure 1 indicates expression in humans in the uterus, lung, germ cells, liver, parathyroid gland, prostate, and placenta.

FIGURE 2 provides the predicted amino acid sequence of the transporter of the present invention. (SEQ ID NO:2) In addition structure and functional information such as protein family, function, and modification sites is provided where available, allowing one to readily determine specific uses of inventions based on this molecular sequence.

FIGURE 3 provides genomic sequences that span the gene encoding the transporter protein of the present invention. (SEQ ID NO:3) In addition structure and functional information, such as intron/exon structure, promoter location, etc., is provided where available, allowing one to readily determine specific uses of inventions based on this molecular sequence. As illustrated in Figure 3, SNPs were identified at 19 different nucleotide positions.

DETAILED DESCRIPTION OF THE INVENTION

General Description

The present invention is based on the sequencing of the human genome. During the sequencing and assembly of the human genome, analysis of the sequence information revealed previously unidentified fragments of the human genome that encode peptides that share structural and/or sequence homology to protein/peptide/domains identified and characterized within the art as being a transporter protein or part of a transporter protein and are related to the ion channel transporter subfamily. Utilizing these sequences, additional genomic sequences were assembled and transcript and/or cDNA sequences were isolated and characterized. Based on this analysis, the present invention provides amino acid sequences of human transporter peptides and proteins that are related to the ion channel transporter subfamily, nucleic acid sequences in the form of transcript sequences, cDNA sequences and/or genomic sequences that encode these transporter peptides and proteins, nucleic acid variation (allelic information), tissue distribution of expression, and information about the closest art known protein/peptide/domain that has structural or sequence homology to the transporter of the present invention.

In addition to being previously unknown, the peptides that are provided in the present invention are selected based on their ability to be used for the development of commercially important products and services. Specifically, the present peptides are selected based on homology and/or structural relatedness to known transporter proteins of the ion channel transporter subfamily and the expression pattern observed. Experimental data as provided in Figure 1 indicates expression in humans in the uterus, lung, germ cells, liver, parathyroid gland, prostate, and placenta.. The art has clearly established the commercial importance of members of this family of proteins and proteins that have expression patterns similar to that of the present gene. Some of the more specific features of the peptides of the present invention, and the uses thereof, are described herein, particularly in the Background of the Invention and in the annotation provided in the Figures, and/or are known within the art for each of the known ion channel family or subfamily of transporter proteins.

Specific Embodiments

Peptide Molecules

The present invention provides nucleic acid sequences that encode protein molecules that have been identified as being members of the transporter family of proteins and are related to the ion channel transporter subfamily (protein sequences are provided in Figure 2, transcript/cDNA sequences are provided in Figures 1 and genomic sequences are provided in Figure 3). The peptide sequences provided in Figure 2, as well as the obvious variants described herein, particularly allelic variants as identified herein and using the information in Figure 3, will be referred herein as the transporter peptides of the present invention, transporter peptides, or peptides/proteins of the present invention.

The present invention provides isolated peptide and protein molecules that consist of, consist essentially of, or comprising the amino acid sequences of the transporter peptides disclosed in the Figure 2, (encoded by the nucleic acid molecule shown in Figure 1, transcript/cDNA or Figure 3, genomic sequence), as well as all obvious variants

of these peptides that are within the art to make and use. Some of these variants are described in detail below.

As used herein, a peptide is said to be "isolated" or "purified" when it is substantially free of cellular material or free of chemical precursors or other chemicals. The peptides of the present invention can be purified to homogeneity or other degrees of purity. The level of purification will be based on the intended use. The critical feature is that the preparation allows for the desired function of the peptide, even if in the presence of considerable amounts of other components (the features of an isolated nucleic acid molecule is discussed below).

In some uses, "substantially free of cellular material" includes preparations of the peptide having less than about 30% (by dry weight) other proteins (i.e., contaminating protein), less than about 20% other proteins, less than about 10% other proteins, or less than about 5% other proteins. When the peptide is recombinantly produced, it can also be substantially free of culture medium, i.e., culture medium represents less than about 20% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of the peptide in which it is separated from chemical precursors or other chemicals that are involved in its synthesis. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of the transporter peptide having less than about 30% (by dry weight) chemical precursors or other chemicals, less than about 20% chemical precursors or other chemicals, less than about 10% chemical precursors or other chemicals, or less than about 5% chemical precursors or other chemicals.

The isolated transporter peptide can be purified from cells that naturally express it, purified from cells that have been altered to express it (recombinant), or synthesized using known protein synthesis methods. Experimental data as provided in Figure 1 indicates expression in humans in the uterus, lung, germ cells, liver, parathyroid gland, prostate, and placenta. For example, a nucleic acid molecule encoding the transporter peptide is cloned into an expression vector, the expression vector introduced into a host cell and the protein expressed in the host cell. The protein can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques. Many of these techniques are described in detail below.

Accordingly, the present invention provides proteins that consist of the amino acid sequences provided in Figure 2 (SEQ ID NO:2), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NO:1) and the genomic sequences provided in Figure 3 (SEQ ID NO:3). The amino acid sequence of such a protein is provided in Figure 2. A protein consists of an amino acid sequence when the amino acid sequence is the final amino acid sequence of the protein.

The present invention further provides proteins that consist essentially of the amino acid sequences provided in Figure 2 (SEQ ID NO:2), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NO:1) and the genomic sequences provided in Figure 3 (SEQ ID NO:3). A protein consists essentially of an amino acid sequence when such an amino acid sequence is present with only a few additional amino acid residues, for example from about 1 to about 100 or so additional residues, typically from 1 to about 20 additional residues in the final protein.

The present invention further provides proteins that comprise the amino acid sequences provided in Figure 2 (SEQ ID NO:2), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NO:1) and the genomic sequences provided in Figure 3 (SEQ ID NO:3). A protein comprises an amino acid sequence when the amino acid sequence is at least part of the final amino acid sequence of the protein. In such a fashion, the protein can be only the peptide or have additional amino acid molecules, such as amino acid residues (contiguous encoded sequence) that are naturally associated with it or heterologous amino acid residues/peptide sequences. Such a protein can have a few additional amino acid residues or can comprise several hundred or more additional amino acids. The preferred classes of proteins that are comprised of the transporter peptides of the present invention are the naturally occurring mature proteins. A brief description of how various types of these proteins can be made/isolated is provided below.

The transporter peptides of the present invention can be attached to heterologous sequences to form chimeric or fusion proteins. Such chimeric and fusion proteins comprise a transporter peptide operatively linked to a heterologous protein having an amino acid sequence not substantially homologous to the transporter peptide. "Operatively linked" indicates that the transporter peptide and the heterologous protein are fused in-frame. The

heterologous protein can be fused to the N-terminus or C-terminus of the transporter peptide.

In some uses, the fusion protein does not affect the activity of the transporter peptide *per se*. For example, the fusion protein can include, but is not limited to, enzymatic fusion proteins, for example beta-galactosidase fusions, yeast two-hybrid GAL fusions, poly-His fusions, MYC-tagged, HI-tagged and Ig fusions. Such fusion proteins, particularly poly-His fusions, can facilitate the purification of recombinant transporter peptide. In certain host cells (e.g., mammalian host cells), expression and/or secretion of a protein can be increased by using a heterologous signal sequence.

A chimeric or fusion protein can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different protein sequences are ligated together in-frame in accordance with conventional techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and re-amplified to generate a chimeric gene sequence (see Ausubel *et al.*, *Current Protocols in Molecular Biology*, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST protein). A transporter peptide-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the transporter peptide.

As mentioned above, the present invention also provides and enables obvious variants of the amino acid sequence of the proteins of the present invention, such as naturally occurring mature forms of the peptide, allelic/sequence variants of the peptides, non-naturally occurring recombinantly derived variants of the peptides, and orthologs and paralogs of the peptides. Such variants can readily be generated using art-known techniques in the fields of recombinant nucleic acid technology and protein biochemistry. It is understood, however, that variants exclude any amino acid sequences disclosed prior to the invention.

Such variants can readily be identified/made using molecular techniques and the sequence information disclosed herein. Further, such variants can readily be distinguished

from other peptides based on sequence and/or structural homology to the transporter peptides of the present invention. The degree of homology/identity present will be based primarily on whether the peptide is a functional variant or non-functional variant, the amount of divergence present in the paralog family and the evolutionary distance between the orthologs.

To determine the percent identity of two amino acid sequences or two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, at least 30%, 40%, 50%, 60%, 70%, 80%, or 90% or more of a reference sequence is aligned for comparison purposes. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity and similarity between two sequences can be accomplished using a mathematical algorithm. (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part 1*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (*J. Mol. Biol.* (48):444-453 (1970)) algorithm which has been incorporated into the GAP program in the GCG software package (available at http://www.gcg.com), using either a Blossom 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12,

10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (Devereux, J., *et al., Nucleic Acids Res. 12(1)*:387 (1984)) (available at http://www.gcg.com), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of E. Myers and W. Miller (CABIOS, 4:11-17 (1989)) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

The nucleic acid and protein sequences of the present invention can further be used as a "query sequence" to perform a search against sequence databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, *et al.* (*J. Mol. Biol.* 215:403-10 (1990)). BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to the nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to the proteins of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.* (*Nucleic Acids Res.* 25(17):3389-3402 (1997)). When utilizing BLAST and gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used.

Full-length pre-processed forms, as well as mature processed forms, of proteins that comprise one of the peptides of the present invention can readily be identified as having complete sequence identity to one of the transporter peptides of the present invention as well as being encoded by the same genetic locus as the transporter peptide provided herein. The gene encoding the novel transporter protein of the present invention is located on a genome component that has been mapped to human chromosome X (as indicated in Figure 3), which is supported by multiple lines of evidence, such as STS and BAC map data.

Allelic variants of a transporter peptide can readily be identified as being a human protein having a high degree (significant) of sequence homology/identity to at least a portion

of the transporter peptide as well as being encoded by the same genetic locus as the transporter peptide provided herein. Genetic locus can readily be determined based on the genomic information provided in Figure 3, such as the genomic sequence mapped to the reference human. The gene encoding the novel transporter protein of the present invention is located on a genome component that has been mapped to human chromosome X (as indicated in Figure 3), which is supported by multiple lines of evidence, such as STS and BAC map data. As used herein, two proteins (or a region of the proteins) have significant homology when the amino acid sequences are typically at least about 70-80%, 80-90%, and more typically at least about 90-95% or more homologous. A significantly homologous amino acid sequence, according to the present invention, will be encoded by a nucleic acid sequence that will hybridize to a transporter peptide encoding nucleic acid molecule under stringent conditions as more fully described below.

Figure 3 provides information on SNPs that have been found in the gene encoding the transporter protein of the present invention. SNPs were identified at 19 different nucleotide positions. Some of these SNPs that are located outside the ORF and in introns may affect gene transcription.

Paralogs of a transporter peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the transporter peptide, as being encoded by a gene from humans, and as having similar activity or function. Two proteins will typically be considered paralogs when the amino acid sequences are typically at least about 60% or greater, and more typically at least about 70% or greater homology through a given region or domain. Such paralogs will be encoded by a nucleic acid sequence that will hybridize to a transporter peptide encoding nucleic acid molecule under moderate to stringent conditions as more fully described below.

Orthologs of a transporter peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the transporter peptide as well as being encoded by a gene from another organism. Preferred orthologs will be isolated from mammals, preferably primates, for the development of human therapeutic targets and agents. Such orthologs will be encoded by a nucleic acid sequence that will hybridize to a transporter peptide encoding nucleic acid molecule under moderate to stringent

conditions, as more fully described below, depending on the degree of relatedness of the two organisms yielding the proteins.

Non-naturally occurring variants of the transporter peptides of the present invention can readily be generated using recombinant techniques. Such variants include, but are not limited to deletions, additions and substitutions in the amino acid sequence of the transporter peptide. For example, one class of substitutions are conserved amino acid substitution. Such substitutions are those that substitute a given amino acid in a transporter peptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu, and Ile; interchange of the hydroxyl residues Ser and Thr; exchange of the acidic residues Asp and Glu; substitution between the amide residues Asn and Gln; exchange of the basic residues Lys and Arg; and replacements among the aromatic residues Phe and Tyr. Guidance concerning which amino acid changes are likely to be phenotypically silent are found in Bowie *et al.*, *Science 247*:1306-1310 (1990).

Variant transporter peptides can be fully functional or can lack function in one or more activities, e.g. ability to bind ligand, ability to transport ligand, ability to mediate signaling, etc. Fully functional variants typically contain only conservative variation or variation in non-critical residues or in non-critical regions. Figure 2 provides the result of protein analysis and can be used to identify critical domains/regions. Functional variants can also contain substitution of similar amino acids that result in no change or an insignificant change in function. Alternatively, such substitutions may positively or negatively affect function to some degree.

Non-functional variants typically contain one or more non-conservative amino acid substitutions, deletions, insertions, inversions, or truncation or a substitution, insertion, inversion, or deletion in a critical residue or critical region.

Amino acids that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham *et al.*, *Science 244*:1081-1085 (1989)), particularly using the results provided in Figure 2. The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity such as transporter activity or in assays such as an *in vitro* proliferative activity. Sites that are critical for binding

partner/substrate binding can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling (Smith *et al., J. Mol. Biol. 224*:899-904 (1992); de Vos *et al. Science 255*:306-312 (1992)).

The present invention further provides fragments of the transporter peptides, in addition to proteins and peptides that comprise and consist of such fragments, particularly those comprising the residues identified in Figure 2. The fragments to which the invention pertains, however, are not to be construed as encompassing fragments that may be disclosed publicly prior to the present invention.

As used herein, a fragment comprises at least 8, 10, 12, 14, 16, or more contiguous amino acid residues from a transporter peptide. Such fragments can be chosen based on the ability to retain one or more of the biological activities of the transporter peptide or could be chosen for the ability to perform a function, e.g. bind a substrate or act as an immunogen. Particularly important fragments are biologically active fragments, peptides that are, for example, about 8 or more amino acids in length. Such fragments will typically comprise a domain or motif of the transporter peptide, e.g., active site, a transmembrane domain or a substrate-binding domain. Further, possible fragments include, but are not limited to, domain or motif containing fragments, soluble peptide fragments, and fragments containing immunogenic structures. Predicted domains and functional sites are readily identifiable by computer programs well known and readily available to those of skill in the art (e.g., PROSITE analysis). The results of one such analysis are provided in Figure 2.

Polypeptides often contain amino acids other than the 20 amino acids commonly referred to as the 20 naturally occurring amino acids. Further, many amino acids, including the terminal amino acids, may be modified by natural processes, such as processing and other post-translational modifications, or by chemical modification techniques well known in the art. Common modifications that occur naturally in transporter peptides are described in basic texts, detailed monographs, and the research literature, and they are well known to those of skill in the art (some of these features are identified in Figure 2).

Known modifications include, but are not limited to, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking,

cyclization, disulfide bond formation, demethylation, formation of covalent crosslinks, formation of cystine, formation of pyroglutamate, formylation, gamma carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

Such modifications are well known to those of skill in the art and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as *Proteins - Structure and Molecular Properties*, 2nd Ed., T.E. Creighton, W. H. Freeman and Company, New York (1993). Many detailed reviews are available on this subject, such as by Wold, F., *Posttranslational Covalent Modification of Proteins*, B.C. Johnson, Ed., Academic Press, New York 1-12 (1983); Seifter *et al.* (*Meth. Enzymol. 182*: 626-646 (1990)) and Rattan *et al.* (*Ann. N.Y. Acad. Sci. 663*:48-62 (1992)).

Accordingly, the transporter peptides of the present invention also encompass derivatives or analogs in which a substituted amino acid residue is not one encoded by the genetic code, in which a substituent group is included, in which the mature transporter peptide is fused with another compound, such as a compound to increase the half-life of the transporter peptide (for example, polyethylene glycol), or in which the additional amino acids are fused to the mature transporter peptide, such as a leader or secretory sequence or a sequence for purification of the mature transporter peptide or a pro-protein sequence.

Protein/Peptide Uses

The proteins of the present invention can be used in substantial and specific assays related to the functional information provided in the Figures; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its binding partner or ligand) in biological fluids; and as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state). Where the protein binds or

potentially binds to another protein or ligand (such as, for example, in a transporter-effector protein interaction or transporter-ligand interaction), the protein can be used to identify the binding partner/ligand so as to develop a system to identify inhibitors of the binding interaction. Any or all of these uses are capable of being developed into reagent grade or kit format for commercialization as commercial products.

Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include "Molecular Cloning: A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E. F. Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology: Guide to Molecular Cloning Techniques", Academic Press, Berger, S. L. and A. R. Kimmel eds., 1987.

The potential uses of the peptides of the present invention are based primarily on the source of the protein as well as the class/action of the protein. For example, transporters isolated from humans and their human/mammalian orthologs serve as targets for identifying agents for use in mammalian therapeutic applications, e.g. a human drug, particularly in modulating a biological or pathological response in a cell or tissue that expresses the transporter. Experimental data as provided in Figure 1 indicates that the transporter proteins of the present invention are expressed in humans in the uterus, lung, germ cells, liver, parathyroid gland, prostate, and placenta, as indicated by virtual northern blot analysis. PCR-based tissue screening panels also indicate expression in the liver. A large percentage of pharmaceutical agents are being developed that modulate the activity of transporter proteins, particularly members of the ion channel subfamily (see Background of the Invention). The structural and functional information provided in the Background and Figures provide specific and substantial uses for the molecules of the present invention, particularly in combination with the expression information provided in Figure 1. Experimental data as provided in Figure 1 indicates expression in humans in the uterus, lung, germ cells, liver, parathyroid gland, prostate, and placenta. Such uses can readily be determined using the information provided herein, that known in the art and routine experimentation.

The proteins of the present invention (including variants and fragments that may have been disclosed prior to the present invention) are useful for biological assays related to transporters that are related to members of the ion channel subfamily. Such assays involve

any of the known transporter functions or activities or properties useful for diagnosis and treatment of transporter-related conditions that are specific for the subfamily of transporters that the one of the present invention belongs to, particularly in cells and tissues that express the transporter. Experimental data as provided in Figure 1 indicates that the transporter proteins of the present invention are expressed in humans in the uterus, lung, germ cells, liver, parathyroid gland, prostate, and placenta, as indicated by virtual northern blot analysis. PCR-based tissue screening panels also indicate expression in the liver. The proteins of the present invention are also useful in drug screening assays, in cell-based or cell-free systems ((Hodgson, Bio/technology, 1992, Sept 10(9);973-80). Cell-based systems can be native, i.e., cells that normally express the transporter, as a biopsy or expanded in cell culture. Experimental data as provided in Figure 1 indicates expression in humans in the uterus, lung, germ cells, liver, parathyroid gland, prostate, and placenta. In an alternate embodiment, cell-based assays involve recombinant host cells expressing the transporter protein.

The polypeptides can be used to identify compounds that modulate transporter activity of the protein in its natural state or an altered form that causes a specific disease or pathology associated with the transporter. Both the transporters of the present invention and appropriate variants and fragments can be used in high-throughput screens to assay candidate compounds for the ability to bind to the transporter. These compounds can be further screened against a functional transporter to determine the effect of the compound on the transporter activity. Further, these compounds can be tested in animal or invertebrate systems to determine activity/effectiveness. Compounds can be identified that activate (agonist) or inactivate (antagonist) the transporter to a desired degree.

Further, the proteins of the present invention can be used to screen a compound for the ability to stimulate or inhibit interaction between the transporter protein and a molecule that normally interacts with the transporter protein, e.g. a substrate or a component of the signal pathway that the transporter protein normally interacts (for example, another transporter). Such assays typically include the steps of combining the transporter protein with a candidate compound under conditions that allow the transporter protein, or fragment, to interact with the target molecule, and to detect the formation of a complex between the protein and the target or to detect the biochemical consequence of the interaction with the

transporter protein and the target, such as any of the associated effects of signal transduction such as changes in membrane potential, protein phosphorylation, cAMP turnover, and adenylate cyclase activation, etc.

Candidate compounds include, for example, 1) peptides such as soluble peptides, including Ig-tailed fusion peptides and members of random peptide libraries (see, e.g., Lam *et al.*, *Nature 354*:82-84 (1991); Houghten *et al.*, *Nature 354*:84-86 (1991)) and combinatorial chemistry-derived molecular libraries made of D- and/or L- configuration amino acids; 2) phosphopeptides (e.g., members of random and partially degenerate, directed phosphopeptide libraries, see, e.g., Songyang *et al.*, *Cell 72*:767-778 (1993)); 3) antibodies (e.g., polyclonal, monoclonal, humanized, anti-idiotypic, chimeric, and single chain antibodies as well as Fab, $F(ab')_2$, Fab expression library fragments, and epitope-binding fragments of antibodies); and 4) small organic and inorganic molecules (e.g., molecules obtained from combinatorial and natural product libraries).

One candidate compound is a soluble fragment of the receptor that competes for ligand binding. Other candidate compounds include mutant transporters or appropriate fragments containing mutations that affect transporter function and thus compete for ligand. Accordingly, a fragment that competes for ligand, for example with a higher affinity, or a fragment that binds ligand but does not allow release, is encompassed by the invention.

The invention further includes other end point assays to identify compounds that modulate (stimulate or inhibit) transporter activity. The assays typically involve an assay of events in the signal transduction pathway that indicate transporter activity. Thus, the transport of a ligand, change in cell membrane potential, activation of a protein, a change in the expression of genes that are up- or down-regulated in response to the transporter protein dependent signal cascade can be assayed.

Any of the biological or biochemical functions mediated by the transporter can be used as an endpoint assay. These include all of the biochemical or biochemical/biological events described herein, in the references cited herein, incorporated by reference for these endpoint assay targets, and other functions known to those of ordinary skill in the art or that can be readily identified using the information provided in the Figures, particularly Figure 2. Specifically, a biological function of a cell or tissues that expresses the transporter can be assayed. Experimental data as provided in Figure 1 indicates that the transporter proteins

of the present invention are expressed in humans in the uterus, lung, germ cells, liver, parathyroid gland, prostate, and placenta, as indicated by virtual northern blot analysis. PCR-based tissue screening panels also indicate expression in the liver.

Binding and/or activating compounds can also be screened by using chimeric transporter proteins in which the amino terminal extracellular domain, or parts thereof, the entire transmembrane domain or subregions, such as any of the seven transmembrane segments or any of the intracellular or extracellular loops and the carboxy terminal intracellular domain, or parts thereof, can be replaced by heterologous domains or subregions. For example, a ligand-binding region can be used that interacts with a different ligand then that which is recognized by the native transporter. Accordingly, a different set of signal transduction components is available as an end-point assay for activation. This allows for assays to be performed in other than the specific host cell from which the transporter is derived.

The proteins of the present invention are also useful in competition binding assays in methods designed to discover compounds that interact with the transporter (e.g. binding partners and/or ligands). Thus, a compound is exposed to a transporter polypeptide under conditions that allow the compound to bind or to otherwise interact with the polypeptide. Soluble transporter polypeptide is also added to the mixture. If the test compound interacts with the soluble transporter polypeptide, it decreases the amount of complex formed or activity from the transporter target. This type of assay is particularly useful in cases in which compounds are sought that interact with specific regions of the transporter. Thus, the soluble polypeptide that competes with the target transporter region is designed to contain peptide sequences corresponding to the region of interest.

To perform cell free drug screening assays, it is sometimes desirable to immobilize either the transporter protein, or fragment, or its target molecule to facilitate separation of complexes from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay.

Techniques for immobilizing proteins on matrices can be used in the drug screening assays. In one embodiment, a fusion protein can be provided which adds a domain that allows the protein to be bound to a matrix. For example, glutathione-S-transferase fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis,

MO) or glutathione derivatized microtitre plates, which are then combined with the cell lysates (e.g., ^{35}S-labeled) and the candidate compound, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads are washed to remove any unbound label, and the matrix immobilized and radiolabel determined directly, or in the supernatant after the complexes are dissociated. Alternatively, the complexes can be dissociated from the matrix, separated by SDS-PAGE, and the level of transporter-binding protein found in the bead fraction quantitated from the gel using standard electrophoretic techniques. For example, either the polypeptide or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin using techniques well known in the art. Alternatively, antibodies reactive with the protein but which do not interfere with binding of the protein to its target molecule can be derivatized to the wells of the plate, and the protein trapped in the wells by antibody conjugation. Preparations of a transporter-binding protein and a candidate compound are incubated in the transporter protein-presenting wells and the amount of complex trapped in the well can be quantitated. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the transporter protein target molecule, or which are reactive with transporter protein and compete with the target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the target molecule.

Agents that modulate one of the transporters of the present invention can be identified using one or more of the above assays, alone or in combination. It is generally preferable to use a cell-based or cell free system first and then confirm activity in an animal or other model system. Such model systems are well known in the art and can readily be employed in this context.

Modulators of transporter protein activity identified according to these drug screening assays can be used to treat a subject with a disorder mediated by the transporter pathway, by treating cells or tissues that express the transporter. Experimental data as provided in Figure 1 indicates expression in humans in the uterus, lung, germ cells, liver, parathyroid gland, prostate, and placenta. These methods of treatment include the steps of

administering a modulator of transporter activity in a pharmaceutical composition to a subject in need of such treatment, the modulator being identified as described herein.

In yet another aspect of the invention, the transporter proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos *et al.* (1993) *Cell* 72:223-232; Madura *et al.* (1993) *J. Biol. Chem.* 268:12046-12054; Bartel *et al.* (1993) *Biotechniques* 14:920-924; Iwabuchi *et al.* (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300), to identify other proteins, which bind to or interact with the transporter and are involved in transporter activity. Such transporter-binding proteins are also likely to be involved in the propagation of signals by the transporter proteins or transporter targets as, for example, downstream elements of a transporter-mediated signaling pathway. Alternatively, such transporter-binding proteins are likely to be transporter inhibitors.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a transporter protein is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo,* forming a transporter-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the transporter protein.

This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein in an appropriate animal model. For example, an agent identified as described herein (e.g., a transporter-modulating agent, an antisense transporter nucleic acid molecule, a transporter-specific antibody, or a transporter-

binding partner) can be used in an animal or other model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal or other model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein.

The transporter proteins of the present invention are also useful to provide a target for diagnosing a disease or predisposition to disease mediated by the peptide. Accordingly, the invention provides methods for detecting the presence, or levels of, the protein (or encoding mRNA) in a cell, tissue, or organism. Experimental data as provided in Figure 1 indicates expression in humans in the uterus, lung, germ cells, liver, parathyroid gland, prostate, and placenta. The method involves contacting a biological sample with a compound capable of interacting with the transporter protein such that the interaction can be detected. Such an assay can be provided in a single detection format or a multi-detection format such as an antibody chip array.

One agent for detecting a protein in a sample is an antibody capable of selectively binding to protein. A biological sample includes tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject.

The peptides of the present invention also provide targets for diagnosing active protein activity, disease, or predisposition to disease, in a patient having a variant peptide, particularly activities and conditions that are known for other members of the family of proteins to which the present one belongs. Thus, the peptide can be isolated from a biological sample and assayed for the presence of a genetic mutation that results in aberrant peptide. This includes amino acid substitution, deletion, insertion, rearrangement, (as the result of aberrant splicing events), and inappropriate post-translational modification. Analytic methods include altered electrophoretic mobility, altered tryptic peptide digest, altered transporter activity in cell-based or cell-free assay, alteration in ligand or antibody-binding pattern, altered isoelectric point, direct amino acid sequencing, and any other of the known assay techniques useful for detecting mutations in a protein. Such an assay can be provided in a single detection format or a multi-detection format such as an antibody chip array.

In vitro techniques for detection of peptide include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence using a detection reagent, such as an antibody or protein binding agent. Alternatively, the peptide can be detected *in vivo* in a subject by introducing into the subject a labeled anti-peptide antibody or other types of detection agent. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques. Particularly useful are methods that detect the allelic variant of a peptide expressed in a subject and methods which detect fragments of a peptide in a sample.

The peptides are also useful in pharmacogenomic analysis. Pharmacogenomics deal with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, e.g., Eichelbaum, M. (*Clin. Exp. Pharmacol. Physiol.* 23(10-11):983-985 (1996)), and Linder, M.W. (*Clin. Chem.* 43(2):254-266 (1997)). The clinical outcomes of these variations result in severe toxicity of therapeutic drugs in certain individuals or therapeutic failure of drugs in certain individuals as a result of individual variation in metabolism. Thus, the genotype of the individual can determine the way a therapeutic compound acts on the body or the way the body metabolizes the compound. Further, the activity of drug metabolizing enzymes effects both the intensity and duration of drug action. Thus, the pharmacogenomics of the individual permit the selection of effective compounds and effective dosages of such compounds for prophylactic or therapeutic treatment based on the individual's genotype. The discovery of genetic polymorphisms in some drug metabolizing enzymes has explained why some patients do not obtain the expected drug effects, show an exaggerated drug effect, or experience serious toxicity from standard drug dosages. Polymorphisms can be expressed in the phenotype of the extensive metabolizer and the phenotype of the poor metabolizer. Accordingly, genetic polymorphism may lead to allelic protein variants of the transporter protein in which one or more of the transporter functions in one population is different from those in another population. The peptides thus allow a target to ascertain a genetic predisposition that can affect treatment modality. Thus, in a ligand-based treatment, polymorphism may give rise to amino terminal extracellular domains and/or other ligand-binding regions that are more or less active in ligand binding, and transporter activation. Accordingly, ligand dosage would necessarily be modified to maximize the therapeutic

effect within a given population containing a polymorphism. As an alternative to genotyping, specific polymorphic peptides could be identified.

The peptides are also useful for treating a disorder characterized by an absence of, inappropriate, or unwanted expression of the protein. Experimental data as provided in Figure 1 indicates expression in humans in the uterus, lung, germ cells, liver, parathyroid gland, prostate, and placenta. Accordingly, methods for treatment include the use of the transporter protein or fragments.

Antibodies

The invention also provides antibodies that selectively bind to one of the peptides of the present invention, a protein comprising such a peptide, as well as variants and fragments thereof. As used herein, an antibody selectively binds a target peptide when it binds the target peptide and does not significantly bind to unrelated proteins. An antibody is still considered to selectively bind a peptide even if it also binds to other proteins that are not substantially homologous with the target peptide so long as such proteins share homology with a fragment or domain of the peptide target of the antibody. In this case, it would be understood that antibody binding to the peptide is still selective despite some degree of cross-reactivity.

As used herein, an antibody is defined in terms consistent with that recognized within the art: they are multi-subunit proteins produced by a mammalian organism in response to an antigen challenge. The antibodies of the present invention include polyclonal antibodies and monoclonal antibodies, as well as fragments of such antibodies, including, but not limited to, Fab or $F(ab')_2$, and Fv fragments.

Many methods are known for generating and/or identifying antibodies to a given target peptide. Several such methods are described by Harlow, Antibodies, Cold Spring Harbor Press, (1989).

In general, to generate antibodies, an isolated peptide is used as an immunogen and is administered to a mammalian organism, such as a rat, rabbit or mouse. The full-length protein, an antigenic peptide fragment or a fusion protein can be used. Particularly important fragments are those covering functional domains, such as the domains identified in Figure 2, and domain of sequence homology or divergence amongst the family, such as

those that can readily be identified using protein alignment methods and as presented in the Figures.

Antibodies are preferably prepared from regions or discrete fragments of the transporter proteins. Antibodies can be prepared from any region of the peptide as described herein. However, preferred regions will include those involved in function/activity and/or transporter/binding partner interaction. Figure 2 can be used to identify particularly important regions while sequence alignment can be used to identify conserved and unique sequence fragments.

An antigenic fragment will typically comprise at least 8 contiguous amino acid residues. The antigenic peptide can comprise, however, at least 10, 12, 14, 16 or more amino acid residues. Such fragments can be selected on a physical property, such as fragments correspond to regions that are located on the surface of the protein, e.g., hydrophilic regions or can be selected based on sequence uniqueness (see Figure 2).

Detection on an antibody of the present invention can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

Antibody Uses

The antibodies can be used to isolate one of the proteins of the present invention by standard techniques, such as affinity chromatography or immunoprecipitation. The antibodies can facilitate the purification of the natural protein from cells and recombinantly produced protein expressed in host cells. In addition, such antibodies are useful to detect the

presence of one of the proteins of the present invention in cells or tissues to determine the pattern of expression of the protein among various tissues in an organism and over the course of normal development. Experimental data as provided in Figure 1 indicates that the transporter proteins of the present invention are expressed in humans in the uterus, lung, germ cells, liver, parathyroid gland, prostate, and placenta, as indicated by virtual northern blot analysis. PCR-based tissue screening panels also indicate expression in the liver. Further, such antibodies can be used to detect protein *in situ*, *in vitro*, or in a cell lysate or supernatant in order to evaluate the abundance and pattern of expression. Also, such antibodies can be used to assess abnormal tissue distribution or abnormal expression during development or progression of a biological condition. Antibody detection of circulating fragments of the full length protein can be used to identify turnover.

Further, the antibodies can be used to assess expression in disease states such as in active stages of the disease or in an individual with a predisposition toward disease related to the protein's function. When a disorder is caused by an inappropriate tissue distribution, developmental expression, level of expression of the protein, or expressed/processed form, the antibody can be prepared against the normal protein. Experimental data as provided in Figure 1 indicates expression in humans in the uterus, lung, germ cells, liver, parathyroid gland, prostate, and placenta. If a disorder is characterized by a specific mutation in the protein, antibodies specific for this mutant protein can be used to assay for the presence of the specific mutant protein.

The antibodies can also be used to assess normal and aberrant subcellular localization of cells in the various tissues in an organism. Experimental data as provided in Figure 1 indicates expression in humans in the uterus, lung, germ cells, liver, parathyroid gland, prostate, and placenta. The diagnostic uses can be applied, not only in genetic testing, but also in monitoring a treatment modality. Accordingly, where treatment is ultimately aimed at correcting expression level or the presence of aberrant sequence and aberrant tissue distribution or developmental expression, antibodies directed against the protein or relevant fragments can be used to monitor therapeutic efficacy.

Additionally, antibodies are useful in pharmacogenomic analysis. Thus, antibodies prepared against polymorphic proteins can be used to identify individuals that require modified treatment modalities. The antibodies are also useful as diagnostic tools as an

immunological marker for aberrant protein analyzed by electrophoretic mobility, isoelectric point, tryptic peptide digest, and other physical assays known to those in the art.

The antibodies are also useful for tissue typing. Experimental data as provided in Figure 1 indicates expression in humans in the uterus, lung, germ cells, liver, parathyroid gland, prostate, and placenta. Thus, where a specific protein has been correlated with expression in a specific tissue, antibodies that are specific for this protein can be used to identify a tissue type.

The antibodies are also useful for inhibiting protein function, for example, blocking the binding of the transporter peptide to a binding partner such as a ligand or protein binding partner. These uses can also be applied in a therapeutic context in which treatment involves inhibiting the protein's function. An antibody can be used, for example, to block binding, thus modulating (agonizing or antagonizing) the peptides activity. Antibodies can be prepared against specific fragments containing sites required for function or against intact protein that is associated with a cell or cell membrane. See Figure 2 for structural information relating to the proteins of the present invention.

The invention also encompasses kits for using antibodies to detect the presence of a protein in a biological sample. The kit can comprise antibodies such as a labeled or labelable antibody and a compound or agent for detecting protein in a biological sample; means for determining the amount of protein in the sample; means for comparing the amount of protein in the sample with a standard; and instructions for use. Such a kit can be supplied to detect a single protein or epitope or can be configured to detect one of a multitude of epitopes, such as in an antibody detection array. Arrays are described in detail below for nucleic acid arrays and similar methods have been developed for antibody arrays.

Nucleic Acid Molecules

The present invention further provides isolated nucleic acid molecules that encode a transporter peptide or protein of the present invention (cDNA, transcript and genomic sequence). Such nucleic acid molecules will consist of, consist essentially of, or comprise a nucleotide sequence that encodes one of the transporter peptides of the present invention, an allelic variant thereof, or an ortholog or paralog thereof.

As used herein, an "isolated" nucleic acid molecule is one that is separated from other nucleic acid present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences that naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. However, there can be some flanking nucleotide sequences, for example up to about 5KB, 4KB, 3KB, 2KB, or 1KB or less, particularly contiguous peptide encoding sequences and peptide encoding sequences within the same gene but separated by introns in the genomic sequence. The important point is that the nucleic acid is isolated from remote and unimportant flanking sequences such that it can be subjected to the specific manipulations described herein such as recombinant expression, preparation of probes and primers, and other uses specific to the nucleic acid sequences.

Moreover, an "isolated" nucleic acid molecule, such as a transcript/cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. However, the nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered isolated.

For example, recombinant DNA molecules contained in a vector are considered isolated. Further examples of isolated DNA molecules include recombinant DNA molecules maintained in heterologous host cells or purified (partially or substantially) DNA molecules in solution. Isolated RNA molecules include *in vivo* or *in vitro* RNA transcripts of the isolated DNA molecules of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

Accordingly, the present invention provides nucleic acid molecules that consist of the nucleotide sequence shown in Figure 1 or 3 (SEQ ID NO:1, transcript sequence and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in Figure 2, SEQ ID NO:2. A nucleic acid molecule consists of a nucleotide sequence when the nucleotide sequence is the complete nucleotide sequence of the nucleic acid molecule.

The present invention further provides nucleic acid molecules that consist essentially of the nucleotide sequence shown in Figure 1 or 3 (SEQ ID NO:1, transcript sequence and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the protein

provided in Figure 2, SEQ ID NO:2. A nucleic acid molecule consists essentially of a nucleotide sequence when such a nucleotide sequence is present with only a few additional nucleic acid residues in the final nucleic acid molecule.

The present invention further provides nucleic acid molecules that comprise the nucleotide sequences shown in Figure 1 or 3 (SEQ ID NO:1, transcript sequence and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in Figure 2, SEQ ID NO:2. A nucleic acid molecule comprises a nucleotide sequence when the nucleotide sequence is at least part of the final nucleotide sequence of the nucleic acid molecule. In such a fashion, the nucleic acid molecule can be only the nucleotide sequence or have additional nucleic acid residues, such as nucleic acid residues that are naturally associated with it or heterologous nucleotide sequences. Such a nucleic acid molecule can have a few additional nucleotides or can comprise several hundred or more additional nucleotides. A brief description of how various types of these nucleic acid molecules can be readily made/isolated is provided below.

In Figures 1 and 3, both coding and non-coding sequences are provided. Because of the source of the present invention, humans genomic sequence (Figure 3) and cDNA/transcript sequences (Figure 1), the nucleic acid molecules in the Figures will contain genomic intronic sequences, 5' and 3' non-coding sequences, gene regulatory regions and non-coding intergenic sequences. In general such sequence features are either noted in Figures 1 and 3 or can readily be identified using computational tools known in the art. As discussed below, some of the non-coding regions, particularly gene regulatory elements such as promoters, are useful for a variety of purposes, e.g. control of heterologous gene expression, target for identifying gene activity modulating compounds, and are particularly claimed as fragments of the genomic sequence provided herein.

The isolated nucleic acid molecules can encode the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature peptide (when the mature form has more than one peptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, facilitate protein trafficking, prolong or shorten protein half-life or facilitate manipulation of a protein for assay or production, among other things. As generally is the case *in situ*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

As mentioned above, the isolated nucleic acid molecules include, but are not limited to, the sequence encoding the transporter peptide alone, the sequence encoding the mature peptide and additional coding sequences, such as a leader or secretory sequence (e.g., a pre-pro or pro-protein sequence), the sequence encoding the mature peptide, with or without the additional coding sequences, plus additional non-coding sequences, for example introns and non-coding 5' and 3' sequences such as transcribed but non-translated sequences that play a role in transcription, mRNA processing (including splicing and polyadenylation signals), ribosome binding and stability of mRNA. In addition, the nucleic acid molecule may be fused to a marker sequence encoding, for example, a peptide that facilitates purification.

Isolated nucleic acid molecules can be in the form of RNA, such as mRNA, or in the form DNA, including cDNA and genomic DNA obtained by cloning or produced by chemical synthetic techniques or by a combination thereof. The nucleic acid, especially DNA, can be double-stranded or single-stranded. Single-stranded nucleic acid can be the coding strand (sense strand) or the non-coding strand (anti-sense strand).

The invention further provides nucleic acid molecules that encode fragments of the peptides of the present invention as well as nucleic acid molecules that encode obvious variants of the transporter proteins of the present invention that are described above. Such nucleic acid molecules may be naturally occurring, such as allelic variants (same locus), paralogs (different locus), and orthologs (different organism), or may be constructed by recombinant DNA methods or by chemical synthesis. Such non-naturally occurring variants may be made by mutagenesis techniques, including those applied to nucleic acid molecules, cells, or organisms. Accordingly, as discussed above, the variants can contain nucleotide substitutions, deletions, inversions and insertions. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions.

The present invention further provides non-coding fragments of the nucleic acid molecules provided in Figures 1 and 3. Preferred non-coding fragments include, but are not limited to, promoter sequences, enhancer sequences, gene modulating sequences and gene termination sequences. Such fragments are useful in controlling heterologous gene expression and in developing screens to identify gene-modulating agents. A promoter can

readily be identified as being 5' to the ATG start site in the genomic sequence provided in Figure 3.

A fragment comprises a contiguous nucleotide sequence greater than 12 or more nucleotides. Further, a fragment could at least 30, 40, 50, 100, 250 or 500 nucleotides in length. The length of the fragment will be based on its intended use. For example, the fragment can encode epitope bearing regions of the peptide, or can be useful as DNA probes and primers. Such fragments can be isolated using the known nucleotide sequence to synthesize an oligonucleotide probe. A labeled probe can then be used to screen a cDNA library, genomic DNA library, or mRNA to isolate nucleic acid corresponding to the coding region. Further, primers can be used in PCR reactions to clone specific regions of gene.

A probe/primer typically comprises substantially a purified oligonucleotide or oligonucleotide pair. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 20, 25, 40, 50 or more consecutive nucleotides.

Orthologs, homologs, and allelic variants can be identified using methods well known in the art. As described in the Peptide Section, these variants comprise a nucleotide sequence encoding a peptide that is typically 60-70%, 70-80%, 80-90%, and more typically at least about 90-95% or more homologous to the nucleotide sequence shown in the Figure sheets or a fragment of this sequence. Such nucleic acid molecules can readily be identified as being able to hybridize under moderate to stringent conditions, to the nucleotide sequence shown in the Figure sheets or a fragment of the sequence. Allelic variants can readily be determined by genetic locus of the encoding gene. The gene encoding the novel transporter protein of the present invention is located on a genome component that has been mapped to human chromosome X (as indicated in Figure 3), which is supported by multiple lines of evidence, such as STS and BAC map data.

Figure 3 provides information on SNPs that have been found in the gene encoding the transporter protein of the present invention. SNPs were identified at 19 different nucleotide positions. Some of these SNPs that are located outside the ORF and in introns may affect gene transcription.

As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences

encoding a peptide at least 60-70% homologous to each other typically remain hybridized to each other. The conditions can be such that sequences at least about 60%, at least about 70%, or at least about 80% or more homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. One example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65C. Examples of moderate to low stringency hybridization conditions are well known in the art.

Nucleic Acid Molecule Uses

The nucleic acid molecules of the present invention are useful for probes, primers, chemical intermediates, and in biological assays. The nucleic acid molecules are useful as a hybridization probe for messenger RNA, transcript/cDNA and genomic DNA to isolate full-length cDNA and genomic clones encoding the peptide described in Figure 2 and to isolate cDNA and genomic clones that correspond to variants (alleles, orthologs, etc.) producing the same or related peptides shown in Figure 2. As illustrated in Figure 3, SNPs were identified at 19 different nucleotide positions.

The probe can correspond to any sequence along the entire length of the nucleic acid molecules provided in the Figures. Accordingly, it could be derived from 5' noncoding regions, the coding region, and 3' noncoding regions. However, as discussed, fragments are not to be construed as encompassing fragments disclosed prior to the present invention.

The nucleic acid molecules are also useful as primers for PCR to amplify any given region of a nucleic acid molecule and are useful to synthesize antisense molecules of desired length and sequence.

The nucleic acid molecules are also useful for constructing recombinant vectors. Such vectors include expression vectors that express a portion of, or all of, the peptide sequences. Vectors also include insertion vectors, used to integrate into another nucleic acid molecule sequence, such as into the cellular genome, to alter *in situ* expression of a gene and/or gene product. For example, an endogenous coding sequence can be replaced via

homologous recombination with all or part of the coding region containing one or more specifically introduced mutations.

The nucleic acid molecules are also useful for expressing antigenic portions of the proteins.

The nucleic acid molecules are also useful as probes for determining the chromosomal positions of the nucleic acid molecules by means of *in situ* hybridization methods. The gene encoding the novel transporter protein of the present invention is located on a genome component that has been mapped to human chromosome X (as indicated in Figure 3), which is supported by multiple lines of evidence, such as STS and BAC map data.

The nucleic acid molecules are also useful in making vectors containing the gene regulatory regions of the nucleic acid molecules of the present invention.

The nucleic acid molecules are also useful for designing ribozymes corresponding to all, or a part, of the mRNA produced from the nucleic acid molecules described herein.

The nucleic acid molecules are also useful for making vectors that express part, or all, of the peptides.

The nucleic acid molecules are also useful for constructing host cells expressing a part, or all, of the nucleic acid molecules and peptides.

The nucleic acid molecules are also useful for constructing transgenic animals expressing all, or a part, of the nucleic acid molecules and peptides.

The nucleic acid molecules are also useful as hybridization probes for determining the presence, level, form and distribution of nucleic acid expression. Experimental data as provided in Figure 1 indicates that the transporter proteins of the present invention are expressed in humans in the uterus, lung, germ cells, liver, parathyroid gland, prostate, and placenta, as indicated by virtual northern blot analysis. PCR-based tissue screening panels also indicate expression in the liver.

Accordingly, the probes can be used to detect the presence of, or to determine levels of, a specific nucleic acid molecule in cells, tissues, and in organisms. The nucleic acid whose level is determined can be DNA or RNA. Accordingly, probes corresponding to the peptides described herein can be used to assess expression and/or gene copy number in a

given cell, tissue, or organism. These uses are relevant for diagnosis of disorders involving an increase or decrease in transporter protein expression relative to normal results.

In vitro techniques for detection of mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detecting DNA include Southern hybridizations and *in situ* hybridization.

Probes can be used as a part of a diagnostic test kit for identifying cells or tissues that express a transporter protein, such as by measuring a level of a transporter-encoding nucleic acid in a sample of cells from a subject e.g., mRNA or genomic DNA, or determining if a transporter gene has been mutated. Experimental data as provided in Figure 1 indicates that the transporter proteins of the present invention are expressed in humans in the uterus, lung, germ cells, liver, parathyroid gland, prostate, and placenta, as indicated by virtual northern blot analysis. PCR-based tissue screening panels also indicate expression in the liver.

Nucleic acid expression assays are useful for drug screening to identify compounds that modulate transporter nucleic acid expression.

The invention thus provides a method for identifying a compound that can be used to treat a disorder associated with nucleic acid expression of the transporter gene, particularly biological and pathological processes that are mediated by the transporter in cells and tissues that express it. Experimental data as provided in Figure 1 indicates expression in humans in the uterus, lung, germ cells, liver, parathyroid gland, prostate, and placenta. The method typically includes assaying the ability of the compound to modulate the expression of the transporter nucleic acid and thus identifying a compound that can be used to treat a disorder characterized by undesired transporter nucleic acid expression. The assays can be performed in cell-based and cell-free systems. Cell-based assays include cells naturally expressing the transporter nucleic acid or recombinant cells genetically engineered to express specific nucleic acid sequences.

The assay for transporter nucleic acid expression can involve direct assay of nucleic acid levels, such as mRNA levels, or on collateral compounds involved in the signal pathway. Further, the expression of genes that are up- or down-regulated in response to the transporter protein signal pathway can also be assayed. In this embodiment the regulatory regions of these genes can be operably linked to a reporter gene such as luciferase.

Thus, modulators of transporter gene expression can be identified in a method wherein a cell is contacted with a candidate compound and the expression of mRNA determined. The level of expression of transporter mRNA in the presence of the candidate compound is compared to the level of expression of transporter mRNA in the absence of the candidate compound. The candidate compound can then be identified as a modulator of nucleic acid expression based on this comparison and be used, for example to treat a disorder characterized by aberrant nucleic acid expression. When expression of mRNA is statistically significantly greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of nucleic acid expression. When nucleic acid expression is statistically significantly less in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of nucleic acid expression.

The invention further provides methods of treatment, with the nucleic acid as a target, using a compound identified through drug screening as a gene modulator to modulate transporter nucleic acid expression in cells and tissues that express the transporter. Experimental data as provided in Figure 1 indicates that the transporter proteins of the present invention are expressed in humans in the uterus, lung, germ cells, liver, parathyroid gland, prostate, and placenta, as indicated by virtual northern blot analysis. PCR-based tissue screening panels also indicate expression in the liver. Modulation includes both up-regulation (i.e. activation or agonization) or down-regulation (suppression or antagonization) or nucleic acid expression.

Alternatively, a modulator for transporter nucleic acid expression can be a small molecule or drug identified using the screening assays described herein as long as the drug or small molecule inhibits the transporter nucleic acid expression in the cells and tissues that express the protein. Experimental data as provided in Figure 1 indicates expression in humans in the uterus, lung, germ cells, liver, parathyroid gland, prostate, and placenta.

The nucleic acid molecules are also useful for monitoring the effectiveness of modulating compounds on the expression or activity of the transporter gene in clinical trials or in a treatment regimen. Thus, the gene expression pattern can serve as a barometer for the continuing effectiveness of treatment with the compound, particularly with compounds to which a patient can develop resistance. The gene expression pattern can also serve as a

marker indicative of a physiological response of the affected cells to the compound. Accordingly, such monitoring would allow either increased administration of the compound or the administration of alternative compounds to which the patient has not become resistant. Similarly, if the level of nucleic acid expression falls below a desirable level, administration of the compound could be commensurately decreased.

The nucleic acid molecules are also useful in diagnostic assays for qualitative changes in transporter nucleic acid expression, and particularly in qualitative changes that lead to pathology. The nucleic acid molecules can be used to detect mutations in transporter genes and gene expression products such as mRNA. The nucleic acid molecules can be used as hybridization probes to detect naturally occurring genetic mutations in the transporter gene and thereby to determine whether a subject with the mutation is at risk for a disorder caused by the mutation. Mutations include deletion, addition, or substitution of one or more nucleotides in the gene, chromosomal rearrangement, such as inversion or transposition, modification of genomic DNA, such as aberrant methylation patterns or changes in gene copy number, such as amplification. Detection of a mutated form of the transporter gene associated with a dysfunction provides a diagnostic tool for an active disease or susceptibility to disease when the disease results from overexpression, underexpression, or altered expression of a transporter protein.

Individuals carrying mutations in the transporter gene can be detected at the nucleic acid level by a variety of techniques. Figure 3 provides information on SNPs that have been found in the gene encoding the transporter protein of the present invention. SNPs were identified at 19 different nucleotide positions. Some of these SNPs that are located outside the ORF and in introns may affect gene transcription. The gene encoding the novel transporter protein of the present invention is located on a genome component that has been mapped to human chromosome X (as indicated in Figure 3), which is supported by multiple lines of evidence, such as STS and BAC map data. Genomic DNA can be analyzed directly or can be amplified by using PCR prior to analysis. RNA or cDNA can be used in the same way. In some uses, detection of the mutation involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g. U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran *et al.*, *Science 241*:1077-1080 (1988); and Nakazawa *et al.*, *PNAS 91*:360-364

(1994)), the latter of which can be particularly useful for detecting point mutations in the gene (see Abravaya *et al.*, *Nucleic Acids Res. 23*:675-682 (1995)). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a gene under conditions such that hybridization and amplification of the gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. Deletions and insertions can be detected by a change in size of the amplified product compared to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to normal RNA or antisense DNA sequences.

Alternatively, mutations in a transporter gene can be directly identified, for example, by alterations in restriction enzyme digestion patterns determined by gel electrophoresis.

Further, sequence-specific ribozymes (U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site. Perfectly matched sequences can be distinguished from mismatched sequences by nuclease cleavage digestion assays or by differences in melting temperature.

Sequence changes at specific locations can also be assessed by nuclease protection assays such as RNase and S1 protection or the chemical cleavage method. Furthermore, sequence differences between a mutant transporter gene and a wild-type gene can be determined by direct DNA sequencing. A variety of automated sequencing procedures can be utilized when performing the diagnostic assays (Naeve, C.W., (1995) *Biotechniques 19*:448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen *et al.*, *Adv. Chromatogr. 36*:127-162 (1996); and Griffin *et al.*, *Appl. Biochem. Biotechnol. 38*:147-159 (1993)).

Other methods for detecting mutations in the gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA duplexes (Myers *et al.*, *Science 230*:1242 (1985)); Cotton *et al.*, *PNAS 85*:4397 (1988); Saleeba *et al.*, *Meth. Enzymol. 217*:286-295 (1992)), electrophoretic mobility of mutant and wild type nucleic acid is compared (Orita *et al.*, *PNAS 86*:2766 (1989); Cotton *et al.*, *Mutat. Res. 285*:125-144 (1993); and Hayashi *et al.*, *Genet. Anal. Tech. Appl. 9*:73-79

(1992)), and movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (Myers *et al.*, *Nature 313*:495 (1985)). Examples of other techniques for detecting point mutations include selective oligonucleotide hybridization, selective amplification, and selective primer extension.

The nucleic acid molecules are also useful for testing an individual for a genotype that while not necessarily causing the disease, nevertheless affects the treatment modality. Thus, the nucleic acid molecules can be used to study the relationship between an individual's genotype and the individual's response to a compound used for treatment (pharmacogenomic relationship). Accordingly, the nucleic acid molecules described herein can be used to assess the mutation content of the transporter gene in an individual in order to select an appropriate compound or dosage regimen for treatment. Figure 3 provides information on SNPs that have been found in the gene encoding the transporter protein of the present invention. SNPs were identified at 19 different nucleotide positions. Some of these SNPs that are located outside the ORF and in introns may affect gene transcription.

Thus nucleic acid molecules displaying genetic variations that affect treatment provide a diagnostic target that can be used to tailor treatment in an individual. Accordingly, the production of recombinant cells and animals containing these polymorphisms allow effective clinical design of treatment compounds and dosage regimens.

The nucleic acid molecules are thus useful as antisense constructs to control transporter gene expression in cells, tissues, and organisms. A DNA antisense nucleic acid molecule is designed to be complementary to a region of the gene involved in transcription, preventing transcription and hence production of transporter protein. An antisense RNA or DNA nucleic acid molecule would hybridize to the mRNA and thus block translation of mRNA into transporter protein.

Alternatively, a class of antisense molecules can be used to inactivate mRNA in order to decrease expression of transporter nucleic acid. Accordingly, these molecules can treat a disorder characterized by abnormal or undesired transporter nucleic acid expression. This technique involves cleavage by means of ribozymes containing nucleotide sequences complementary to one or more regions in the mRNA that attenuate the ability of the mRNA

to be translated. Possible regions include coding regions and particularly coding regions corresponding to the catalytic and other functional activities of the transporter protein, such as ligand binding.

The nucleic acid molecules also provide vectors for gene therapy in patients containing cells that are aberrant in transporter gene expression. Thus, recombinant cells, which include the patient's cells that have been engineered *ex vivo* and returned to the patient, are introduced into an individual where the cells produce the desired transporter protein to treat the individual.

The invention also encompasses kits for detecting the presence of a transporter nucleic acid in a biological sample. Experimental data as provided in Figure 1 indicates that the transporter proteins of the present invention are expressed in humans in the uterus, lung, germ cells, liver, parathyroid gland, prostate, and placenta, as indicated by virtual northern blot analysis. PCR-based tissue screening panels also indicate expression in the liver. For example, the kit can comprise reagents such as a labeled or labelable nucleic acid or agent capable of detecting transporter nucleic acid in a biological sample; means for determining the amount of transporter nucleic acid in the sample; and means for comparing the amount of transporter nucleic acid in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect transporter protein mRNA or DNA.

Nucleic Acid Arrays

The present invention further provides nucleic acid detection kits, such as arrays or microarrays of nucleic acid molecules that are based on the sequence information provided in Figures 1 and 3 (SEQ ID NOS:1 and 3).

As used herein "Arrays" or "Microarrays" refers to an array of distinct polynucleotides or oligonucleotides synthesized on a substrate, such as paper, nylon or other type of membrane, filter, chip, glass slide, or any other suitable solid support. In one embodiment, the microarray is prepared and used according to the methods described in US Patent 5,837,832, Chee *et al.*, PCT application W095/11995 (Chee *et al.*), Lockhart, D. J. *et al.* (1996; Nat. Biotech. 14: 1675-1680) and Schena, M. *et al.* (1996; Proc. Natl. Acad. Sci. 93: 10614-10619), all of which are incorporated herein in their

entirety by reference. In other embodiments, such arrays are produced by the methods described by Brown *et al.*, US Patent No. 5,807,522.

The microarray or detection kit is preferably composed of a large number of unique, single-stranded nucleic acid sequences, usually either synthetic antisense oligonucleotides or fragments of cDNAs, fixed to a solid support. The oligonucleotides are preferably about 6-60 nucleotides in length, more preferably 15-30 nucleotides in length, and most preferably about 20-25 nucleotides in length. For a certain type of microarray or detection kit, it may be preferable to use oligonucleotides that are only 7-20 nucleotides in length. The microarray or detection kit may contain oligonucleotides that cover the known 5', or 3', sequence, sequential oligonucleotides that cover the full length sequence; or unique oligonucleotides selected from particular areas along the length of the sequence. Polynucleotides used in the microarray or detection kit may be oligonucleotides that are specific to a gene or genes of interest.

In order to produce oligonucleotides to a known sequence for a microarray or detection kit, the gene(s) of interest (or an ORF identified from the contigs of the present invention) is typically examined using a computer algorithm which starts at the 5' or at the 3' end of the nucleotide sequence. Typical algorithms will then identify oligomers of defined length that are unique to the gene, have a GC content within a range suitable for hybridization, and lack predicted secondary structure that may interfere with hybridization. In certain situations it may be appropriate to use pairs of oligonucleotides on a microarray or detection kit. The "pairs" will be identical, except for one nucleotide that preferably is located in the center of the sequence. The second oligonucleotide in the pair (mismatched by one) serves as a control. The number of oligonucleotide pairs may range from two to one million. The oligomers are synthesized at designated areas on a substrate using a light-directed chemical process. The substrate may be paper, nylon or other type of membrane, filter, chip, glass slide or any other suitable solid support.

In another aspect, an oligonucleotide may be synthesized on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application W095/251116 (Baldeschweiler *et al.*) which is incorporated herein in its entirety by reference. In another aspect, a "gridded" array analogous to a dot (or slot) blot may be used to arrange and link cDNA fragments or oligonucleotides to the

surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedures. An array, such as those described above, may be produced by hand or by using available devices (slot blot or dot blot apparatus), materials (any suitable solid support), and machines (including robotic instruments), and may contain 8, 24, 96, 384, 1536, 6144 or more oligonucleotides, or any other number between two and one million which lends itself to the efficient use of commercially available instrumentation.

In order to conduct sample analysis using a microarray or detection kit, the RNA or DNA from a biological sample is made into hybridization probes. The mRNA is isolated, and cDNA is produced and used as a template to make antisense RNA (aRNA). The aRNA is amplified in the presence of fluorescent nucleotides, and labeled probes are incubated with the microarray or detection kit so that the probe sequences hybridize to complementary oligonucleotides of the microarray or detection kit. Incubation conditions are adjusted so that hybridization occurs with precise complementary matches or with various degrees of less complementarity. After removal of nonhybridized probes, a scanner is used to determine the levels and patterns of fluorescence. The scanned images are examined to determine degree of complementarity and the relative abundance of each oligonucleotide sequence on the microarray or detection kit. The biological samples may be obtained from any bodily fluids (such as blood, urine, saliva, phlegm, gastric juices, etc.), cultured cells, biopsies, or other tissue preparations. A detection system may be used to measure the absence, presence, and amount of hybridization for all of the distinct sequences simultaneously. This data may be used for large-scale correlation studies on the sequences, expression patterns, mutations, variants, or polymorphisms among samples.

Using such arrays, the present invention provides methods to identify the expression of the transporter proteins/peptides of the present invention. In detail, such methods comprise incubating a test sample with one or more nucleic acid molecules and assaying for binding of the nucleic acid molecule with components within the test sample. Such assays will typically involve arrays comprising many genes, at least one of which is a gene of the present invention and or alleles of the transporter gene of the present invention. Figure 3 provides information on SNPs that have been found in the gene encoding the transporter protein of the present invention. SNPs were identified at 19

different nucleotide positions. Some of these SNPs that are located outside the ORF and in introns may affect gene transcription.

Conditions for incubating a nucleic acid molecule with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the nucleic acid molecule used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or array assay formats can readily be adapted to employ the novel fragments of the Human genome disclosed herein. Examples of such assays can be found in Chard, T, *An Introduction to Radioimmunoassay and Related Techniques,* Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G. R. *et al., Techniques in Immunocytochemistry,* Academic Press, Orlando, FL Vol. 1 (1 982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., *Practice and Theory of Enzyme Immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology,* Elsevier Science Publishers, Amsterdam, The Netherlands (1985).

The test samples of the present invention include cells, protein or membrane extracts of cells. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing nucleic acid extracts or of cells are well known in the art and can be readily be adapted in order to obtain a sample that is compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention.

Specifically, the invention provides a compartmentalized kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the nucleic acid molecules that can bind to a fragment of the Human genome disclosed herein; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound nucleic acid.

In detail, a compartmentalized kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers, strips of plastic, glass or paper, or arraying material such as silica. Such containers allows one to efficiently transfer reagents from one compartment to another compartment

such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the nucleic acid probe, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound probe. One skilled in the art will readily recognize that the previously unidentified transporter gene of the present invention can be routinely identified using the sequence information disclosed herein can be readily incorporated into one of the established kit formats which are well known in the art, particularly expression arrays.

Vectors/host cells

The invention also provides vectors containing the nucleic acid molecules described herein. The term "vector" refers to a vehicle, preferably a nucleic acid molecule, which can transport the nucleic acid molecules. When the vector is a nucleic acid molecule, the nucleic acid molecules are covalently linked to the vector nucleic acid. With this aspect of the invention, the vector includes a plasmid, single or double stranded phage, a single or double stranded RNA or DNA viral vector, or artificial chromosome, such as a BAC, PAC, YAC, OR MAC.

A vector can be maintained in the host cell as an extrachromosomal element where it replicates and produces additional copies of the nucleic acid molecules. Alternatively, the vector may integrate into the host cell genome and produce additional copies of the nucleic acid molecules when the host cell replicates.

The invention provides vectors for the maintenance (cloning vectors) or vectors for expression (expression vectors) of the nucleic acid molecules. The vectors can function in procaryotic or eukaryotic cells or in both (shuttle vectors).

Expression vectors contain cis-acting regulatory regions that are operably linked in the vector to the nucleic acid molecules such that transcription of the nucleic acid molecules is allowed in a host cell. The nucleic acid molecules can be introduced into the host cell with a separate nucleic acid molecule capable of affecting transcription. Thus, the second nucleic acid molecule may provide a trans-acting factor interacting with the cis-regulatory

control region to allow transcription of the nucleic acid molecules from the vector. Alternatively, a trans-acting factor may be supplied by the host cell. Finally, a trans-acting factor can be produced from the vector itself. It is understood, however, that in some embodiments, transcription and/or translation of the nucleic acid molecules can occur in a cell-free system.

The regulatory sequence to which the nucleic acid molecules described herein can be operably linked include promoters for directing mRNA transcription. These include, but are not limited to, the left promoter from bacteriophage λ, the lac, TRP, and TAC promoters from *E. coli*, the early and late promoters from SV40, the CMV immediate early promoter, the adenovirus early and late promoters, and retrovirus long-terminal repeats.

In addition to control regions that promote transcription, expression vectors may also include regions that modulate transcription, such as repressor binding sites and enhancers. Examples include the SV40 enhancer, the cytomegalovirus immediate early enhancer, polyoma enhancer, adenovirus enhancers, and retrovirus LTR enhancers.

In addition to containing sites for transcription initiation and control, expression vectors can also contain sequences necessary for transcription termination and, in the transcribed region a ribosome binding site for translation. Other regulatory control elements for expression include initiation and termination codons as well as polyadenylation signals. The person of ordinary skill in the art would be aware of the numerous regulatory sequences that are useful in expression vectors. Such regulatory sequences are described, for example, in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual. 2nd. ed.*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

A variety of expression vectors can be used to express a nucleic acid molecule. Such vectors include chromosomal, episomal, and virus-derived vectors, for example vectors derived from bacterial plasmids, from bacteriophage, from yeast episomes, from yeast chromosomal elements, including yeast artificial chromosomes, from viruses such as baculoviruses, papovaviruses such as SV40, Vaccinia viruses, adenoviruses, poxviruses, pseudorabies viruses, and retroviruses. Vectors may also be derived from combinations of these sources such as those derived from plasmid and bacteriophage genetic elements, e.g. cosmids and phagemids. Appropriate cloning and expression vectors for prokaryotic and

eukaryotic hosts are described in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual. 2nd. ed.*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

The regulatory sequence may provide constitutive expression in one or more host cells (i.e. tissue specific) or may provide for inducible expression in one or more cell types such as by temperature, nutrient additive, or exogenous factor such as a hormone or other ligand. A variety of vectors providing for constitutive and inducible expression in prokaryotic and eukaryotic hosts are well known to those of ordinary skill in the art.

The nucleic acid molecules can be inserted into the vector nucleic acid by well-known methodology. Generally, the DNA sequence that will ultimately be expressed is joined to an expression vector by cleaving the DNA sequence and the expression vector with one or more restriction enzymes and then ligating the fragments together. Procedures for restriction enzyme digestion and ligation are well known to those of ordinary skill in the art.

The vector containing the appropriate nucleic acid molecule can be introduced into an appropriate host cell for propagation or expression using well-known techniques. Bacterial cells include, but are not limited to, *E. coli*, *Streptomyces*, and *Salmonella typhimurium*. Eukaryotic cells include, but are not limited to, yeast, insect cells such as *Drosophila*, animal cells such as COS and CHO cells, and plant cells.

As described herein, it may be desirable to express the peptide as a fusion protein. Accordingly, the invention provides fusion vectors that allow for the production of the peptides. Fusion vectors can increase the expression of a recombinant protein, increase the solubility of the recombinant protein, and aid in the purification of the protein by acting for example as a ligand for affinity purification. A proteolytic cleavage site may be introduced at the junction of the fusion moiety so that the desired peptide can ultimately be separated from the fusion moiety. Proteolytic enzymes include, but are not limited to, factor Xa, thrombin, and enterotransporter. Typical fusion expression vectors include pGEX (Smith *et al.*, *Gene 67*:31-40 (1988)), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann *et al.*, *Gene 69*:301-

315 (1988)) and pET 11d (Studier *et al.*, *Gene Expression Technology: Methods in Enzymology 185*:60-89 (1990)).

Recombinant protein expression can be maximized in host bacteria by providing a genetic background wherein the host cell has an impaired capacity to proteolytically cleave the recombinant protein. (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Alternatively, the sequence of the nucleic acid molecule of interest can be altered to provide preferential codon usage for a specific host cell, for example *E. coli*. (Wada *et al.*, *Nucleic Acids Res. 20*:2111-2118 (1992)).

The nucleic acid molecules can also be expressed by expression vectors that are operative in yeast. Examples of vectors for expression in yeast e.g., *S. cerevisiae* include pYepSec1 (Baldari, *et al.*, *EMBO J. 6*:229-234 (1987)), pMFa (Kurjan *et al.*, *Cell 30*:933-943(1982)), pJRY88 (Schultz *et al.*, *Gene 54*:113-123 (1987)), and pYES2 (Invitrogen Corporation, San Diego, CA).

The nucleic acid molecules can also be expressed in insect cells using, for example, baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith *et al.*, *Mol. Cell Biol. 3*:2156-2165 (1983)) and the pVL series (Lucklow *et al.*, *Virology 170*:31-39 (1989)).

In certain embodiments of the invention, the nucleic acid molecules described herein are expressed in mammalian cells using mammalian expression vectors. Examples of mammalian expression vectors include pCDM8 (Seed, B. *Nature 329*:840(1987)) and pMT2PC (Kaufman *et al.*, *EMBO J. 6*:187-195 (1987)).

The expression vectors listed herein are provided by way of example only of the well-known vectors available to those of ordinary skill in the art that would be useful to express the nucleic acid molecules. The person of ordinary skill in the art would be aware of other vectors suitable for maintenance propagation or expression of the nucleic acid molecules described herein. These are found for example in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

The invention also encompasses vectors in which the nucleic acid sequences described herein are cloned into the vector in reverse orientation, but operably linked to a

regulatory sequence that permits transcription of antisense RNA. Thus, an antisense transcript can be produced to all, or to a portion, of the nucleic acid molecule sequences described herein, including both coding and non-coding regions. Expression of this antisense RNA is subject to each of the parameters described above in relation to expression of the sense RNA (regulatory sequences, constitutive or inducible expression, tissue-specific expression).

The invention also relates to recombinant host cells containing the vectors described herein. Host cells therefore include prokaryotic cells, lower eukaryotic cells such as yeast, other eukaryotic cells such as insect cells, and higher eukaryotic cells such as mammalian cells.

The recombinant host cells are prepared by introducing the vector constructs described herein into the cells by techniques readily available to the person of ordinary skill in the art. These include, but are not limited to, calcium phosphate transfection, DEAE-dextran-mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, lipofection, and other techniques such as those found in Sambrook, *et al.* (*Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

Host cells can contain more than one vector. Thus, different nucleotide sequences can be introduced on different vectors of the same cell. Similarly, the nucleic acid molecules can be introduced either alone or with other nucleic acid molecules that are not related to the nucleic acid molecules such as those providing trans-acting factors for expression vectors. When more than one vector is introduced into a cell, the vectors can be introduced independently, co-introduced or joined to the nucleic acid molecule vector.

In the case of bacteriophage and viral vectors, these can be introduced into cells as packaged or encapsulated virus by standard procedures for infection and transduction. Viral vectors can be replication-competent or replication-defective. In the case in which viral replication is defective, replication will occur in host cells providing functions that complement the defects.

Vectors generally include selectable markers that enable the selection of the subpopulation of cells that contain the recombinant vector constructs. The marker can be contained in the same vector that contains the nucleic acid molecules described herein or

may be on a separate vector. Markers include tetracycline or ampicillin-resistance genes for prokaryotic host cells and dihydrofolate reductase or neomycin resistance for eukaryotic host cells. However, any marker that provides selection for a phenotypic trait will be effective.

While the mature proteins can be produced in bacteria, yeast, mammalian cells, and other cells under the control of the appropriate regulatory sequences, cell- free transcription and translation systems can also be used to produce these proteins using RNA derived from the DNA constructs described herein.

Where secretion of the peptide is desired, which is difficult to achieve with multi-transmembrane domain containing proteins such as transporters, appropriate secretion signals are incorporated into the vector. The signal sequence can be endogenous to the peptides or heterologous to these peptides.

Where the peptide is not secreted into the medium, which is typically the case with transporters, the protein can be isolated from the host cell by standard disruption procedures, including freeze thaw, sonication, mechanical disruption, use of lysing agents and the like. The peptide can then be recovered and purified by well-known purification methods including ammonium sulfate precipitation, acid extraction, anion or cationic exchange chromatography, phosphocellulose chromatography, hydrophobic-interaction chromatography, affinity chromatography, hydroxylapatite chromatography, lectin chromatography, or high performance liquid chromatography.

It is also understood that depending upon the host cell in recombinant production of the peptides described herein, the peptides can have various glycosylation patterns, depending upon the cell, or maybe non-glycosylated as when produced in bacteria. In addition, the peptides may include an initial modified methionine in some cases as a result of a host-mediated process.

Uses of vectors and host cells

The recombinant host cells expressing the peptides described herein have a variety of uses. First, the cells are useful for producing a transporter protein or peptide that can be further purified to produce desired amounts of transporter protein or fragments. Thus, host cells containing expression vectors are useful for peptide production.

Host cells are also useful for conducting cell-based assays involving the transporter protein or transporter protein fragments, such as those described above as well as other formats known in the art. Thus, a recombinant host cell expressing a native transporter protein is useful for assaying compounds that stimulate or inhibit transporter protein function.

Host cells are also useful for identifying transporter protein mutants in which these functions are affected. If the mutants naturally occur and give rise to a pathology, host cells containing the mutations are useful to assay compounds that have a desired effect on the mutant transporter protein (for example, stimulating or inhibiting function) which may not be indicated by their effect on the native transporter protein.

Genetically engineered host cells can be further used to produce non-human transgenic animals. A transgenic animal is preferably a mammal, for example a rodent, such as a rat or mouse, in which one or more of the cells of the animal include a transgene. A transgene is exogenous DNA that is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal in one or more cell types or tissues of the transgenic animal. These animals are useful for studying the function of a transporter protein and identifying and evaluating modulators of transporter protein activity. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, and amphibians.

A transgenic animal can be produced by introducing nucleic acid into the male pronuclei of a fertilized oocyte, e.g., by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. Any of the transporter protein nucleotide sequences can be introduced as a transgene into the genome of a non-human animal, such as a mouse.

Any of the regulatory or other sequences useful in expression vectors can form part of the transgenic sequence. This includes intronic sequences and polyadenylation signals, if not already included. A tissue-specific regulatory sequence(s) can be operably linked to the transgene to direct expression of the transporter protein to particular cells.

Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder *et*

al., U.S. Patent No. 4,873,191 by Wagner *et al.* and in Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the transgene in its genome and/or expression of transgenic mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene can further be bred to other transgenic animals carrying other transgenes. A transgenic animal also includes animals in which the entire animal or tissues in the animal have been produced using the homologously recombinant host cells described herein.

In another embodiment, transgenic non-human animals can be produced which contain selected systems that allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, see, e.g., Lakso *et al. PNAS 89*:6232-6236 (1992). Another example of a recombinase system is the FLP recombinase system of *S. cerevisiae* (O'Gorman *et al. Science 251*:1351-1355 (1991). If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein is required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, I. *et al. Nature 385*:810-813 (1997) and PCT International Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell, e.g., a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G_0 phase. The quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyst and then transferred to pseudopregnant female foster animal. The offspring born of this female foster animal will be a clone of the animal from which the cell, e.g., the somatic cell, is isolated.

Transgenic animals containing recombinant cells that express the peptides described herein are useful to conduct the assays described herein in an *in vivo* context. Accordingly, the various physiological factors that are present *in vivo* and that could effect ligand binding, transporter protein activation, and signal transduction, may not be evident from *in vitro* cell-free or cell-based assays. Accordingly, it is useful to provide non-human transgenic animals to assay *in vivo* transporter protein function, including ligand interaction, the effect of specific mutant transporter proteins on transporter protein function and ligand interaction, and the effect of chimeric transporter proteins. It is also possible to assess the effect of null mutations, that is mutations that substantially or completely eliminate one or more transporter protein functions.

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the above-described modes for carrying out the invention which are obvious to those skilled in the field of molecular biology or related fields are intended to be within the scope of the following claims.

SEQUENCE LISTING

<110> ZHANG, Hongyu et al.

<120> ISOLATED HUMAN TRANSPORTER PROTEINS,
NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
AND USES THEREOF

<130> CL001208

<160> 4

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 2662
<212> DNA
<213> Human

<400> 1
ctggctccag gtctgactca gtccactaca agctagacgg tcttcttaaa gcaccaacat 60
tacttgagtc tttggataaa attgagaaaa gagtctacaa gtattgtgga ctctacagga 120
ggcaggaggc tgacaactgg cagtaaagac aaagatgtca ggcctgcggc ccggcactca 180
agtggaccct gagattgagc tttttgtaaa ggctggaagt gatggagaga gtattggaaa 240
ctgtcccttt tgccaacgcc ttttcatgat cctctggctt aaaggagtta aatttaatgt 300
gacaactgtt gacatgacca gaaagcctga agaactaaag gacttagccc caggtaccaa 360
tcctccgttc ctggtgtata acaaggagtt gaaaacagac ttcattaaaa ttgaggagtt 420
tttagaacaa accctggctc ctccaaggta ccctcacctg agtcccaagt acaaggagtc 480
ttttgatgtg ggctgtaacc tctttgccaa gttttctgca tacattaaga atacacaaaa 540
ggaggcaaat aagaattttg aaaaatctct gctcaaagaa ttcaagcgtc tggatgacta 600
cttaaacacc ccacttctgg atgaaattga tccagacagt gctgaggaac ccccagtttc 660
cagaagacta ttcttggatg gggaccagct aacactggct gattgtagct tgttacccaa 720
gctgaacatt attaaagttg ctgccaagaa atatcgtgac tttgacattc cagcagaatt 780
ctcaggagtc tggcgttatc tccacaatgc ctatgcccgt gaagaattta cccacacgtg 840
tcctgaagac aaagaaattg aaaatactta cgcaaatgtg gctaaacaga agagttagga 900
gagctcttac aggagaaaag gctatatttg tgatcagatt ttacttattg acatattaga 960
aaggtttttg caaataagaa tatgaaaaat actgtttctt ctatccaact ctcttatgaa 1020
aaggaactct gtattttcta ttagccataa ataatctgtc cactgtattt tacaggtctt 1080
catactttta cttaattttc tttatctgta tggcaaacca ctgcaatcct gaatgacatg 1140
gaaagcatca caatcttttg ccctttgctt gaattcctgg aatgcataca tataagctaa 1200
acagatgtct gcagttataa atgtcataag tagaggtaca atctcaccct gctccttaga 1260
aacatttcca tataaatcgc taaaataatt tcacattttt gttagtttaa tatatacatg 1320
agtttatttc tgatataaat aataaataca gagagtgagc atatcagaga ggcaaattct 1380
taaagaatga tttttaaaat cagctctagg aagagctcaa gatcaattgg tcatagaaca 1440
gcatttgacg cctagaacta tgaccacctc atggtcagag atgagaatgt agcctttgtg 1500
accagattat attattttta aatgaagaag cactcattaa ataaaacata attttaaaaa 1560
acaatataag aaacaaagtc aactgaatct tttattcata gaaatgaaaa ggaaaataaa 1620
aactgtggct gaccaaaagg tcttcttgtt gtccataaaa ggataaggta aacagtcctt 1680
agataattac aaaactttct acaaaagtta aaatgttaca ttactatacg tattcagatt 1740
cacttgttaa agtactctta aatcattcaa atctggaaac aaaagctgaa cttaactctt 1800
gctccctcaa aagagaaaca caagcataag tgcagcttca aaaaaggaaa atattttagg 1860
ctttggtgga agggtggagt ttagataaaa tttaaatgaa gtagcgtttt aataggttca 1920
aagaaaagta aggcaatgag caaactcaaa gtactgtcct tgaaaaccat agagtcaaga 1980
taaatgtata gtgtatggtt aggtggcaga gaaatgcaat catgttgata atctttgaga 2040
tacatcctgt catcagtata tttcagaata catgcaatgc actagcaagt tacaattgat 2100

```
agaatacatt tgaaatgtta aatgaaataa gccaggcaca gaaagacaaa caccacatga 2160
tctcactcat atgtggaatt ttaaaaagtt gatctcactc atatgtggaa ttttaaaaag 2220
ttgatctcac acaagtagag ggtagaatcg tggttaccag gggctaggga gagaaagaag 2280
gcagaggcac tgaaagatgt tggtcaatgg gtataaagtt acacctagga agaataaatt 2340
ttggtattca ccacagtagg gtgactatag caaataataa tgtagcatgt atttcaagat 2400
agctagaaaa gcaggttttt aaatgtcacc acaaagaaat aacaaatgtt tatagtggtg 2460
gatatggtaa ttacgcctat ttgatcatta tactgtgtgt acatgcattg aaacaccaca 2520
ttgtatccca tatatatgta caattatgtg cccattatac atttaaaaaa taaattttaa 2580
aaaccttcaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 2640
aaaaaaaaaa aaaaaaaaaa aa                                          2662
```

<210> 2
<211> 247
<212> PRT
<213> Human

<400> 2

```
Met Ser Gly Leu Arg Pro Gly Thr Gln Val Asp Pro Glu Ile Glu Leu
 1               5                  10                  15
Phe Val Lys Ala Gly Ser Asp Gly Glu Ser Ile Gly Asn Cys Pro Phe
            20                  25                  30
Cys Gln Arg Leu Phe Met Ile Leu Trp Leu Lys Gly Val Lys Phe Asn
        35                  40                  45
Val Thr Thr Val Asp Met Thr Arg Lys Pro Glu Glu Leu Lys Asp Leu
    50                  55                  60
Ala Pro Gly Thr Asn Pro Pro Phe Leu Val Tyr Asn Lys Glu Leu Lys
65                  70                  75                  80
Thr Asp Phe Ile Lys Ile Glu Glu Phe Leu Glu Gln Thr Leu Ala Pro
                85                  90                  95
Pro Arg Tyr Pro His Leu Ser Pro Lys Tyr Lys Glu Ser Phe Asp Val
            100                 105                 110
Gly Cys Asn Leu Phe Ala Lys Phe Ser Ala Tyr Ile Lys Asn Thr Gln
        115                 120                 125
Lys Glu Ala Asn Lys Asn Phe Glu Lys Ser Leu Leu Lys Glu Phe Lys
    130                 135                 140
Arg Leu Asp Asp Tyr Leu Asn Thr Pro Leu Leu Asp Glu Ile Asp Pro
145                 150                 155                 160
Asp Ser Ala Glu Glu Pro Pro Val Ser Arg Arg Leu Phe Leu Asp Gly
                165                 170                 175
Asp Gln Leu Thr Leu Ala Asp Cys Ser Leu Leu Pro Lys Leu Asn Ile
            180                 185                 190
Ile Lys Val Ala Ala Lys Lys Tyr Arg Asp Phe Asp Ile Pro Ala Glu
        195                 200                 205
Phe Ser Gly Val Trp Arg Tyr Leu His Asn Ala Tyr Ala Arg Glu Glu
    210                 215                 220
Phe Thr His Thr Cys Pro Glu Asp Lys Glu Ile Glu Asn Thr Tyr Ala
225                 230                 235                 240
Asn Val Ala Lys Gln Lys Ser
                245
```

<210> 3
<211> 59446
<212> DNA
<213> Human

<220>

<221> misc_feature
<222> (1)...(59446)
<223> n = A,T,C or G

<400> 3
agaactaatc atggttcctg atacagacgc caaaacaagg aagtgatctg ttccagtcca 60
agcttccaag aaataaagaa ctaggtgggg cacactaaac aagcccccag actcaaccac 120
cccagtgaac attccctggt tgtagagaga agtgaaattt gcaacccaga acagaaatct 180
ggctgtgtga gcagtaggat tgggggtgga aacatttaat gaagtacaat tttttaaccc 240
tcttttagac agtatcactg gataaacatc cttttcaata ataaaaatcc aagtcatttc 300
tggccctttt cctggaagtg ctttcaagtt acaggaacac caataagagg ccctttctg 360
ggcatggagc ccaggtctca aagggaggct ctagaaaaca tctggtctgc ttgatatata 420
gaaactagca ctgcatgtgt gtgtttctgt gcatgtgttt ctcctgtgct gactcatggc 480
attgaagcct ctctggaaac acccccaccc ttctagccag gcagtttata cacacccctt 540
tggctcctcc ttgatttaaa tgttagatca cgaggaagaa ggaaaacgat ttcaagagct 600
gcacttaagc atctagaatt ttctgcgtca cacctcttga gagaagagac tggctccagg 660
tctgactcag tccactacaa gctagacggt cttcttaaag caccaacatt acttgagtct 720
ttggataaaa ttgagaaaag agtctacaag tattgtggac tctacaggag gcaggaggct 780
gacaactggc agtaaagaca aagatgtcag gcctgcggcc cggcactcaa gtggaccctg 840
agattgagct ttttgtaaag gtaagttttc cagttataat aactgcatgt agaatatatt 900
agtttttgac actgaagtcc aatgtcttta aaaattctcc acatttgggc tagagatagg 960
aaagaatgtt gtgattattt tcctactctg agttctagaa gaatgcccgg gtgtgtgact 1020
gttcttagat gacaacagga aaacagatct cttctgaaaa aggcaaggtg atatggtgga 1080
aaagcactag actggtttgt agtgagcgac taaattatat tcttaatggc ttcctatata 1140
accttagaaa aatccctcct tctctccaga ctttttttc tccatctata caatgaagga 1200
gcatgacaag atgatcctta agggctttcc aagtctcaaa atctgtgttt tatgagatag 1260
gttttggaag gcctgactgg gtggaggaga gggccgagaa tgacctgaga actccattcc 1320
cacacatagc ctagacagaa ctttctaaac ttctacaatg gacaaacatc acagcagggt 1380
cacatggaca ctgggagaaa aaaaacagga gtctgtgtgc ttgttatgtg aggaggggga 1440
cattttagaa tgctctgctt cttctttttg gtctgccatg gagttgtttt ttttttttt 1500
taacatgtca acttttcaga aaagcacttt ggaaaacccc taaatcaaga gaaaggaaca 1560
tgtgtttcca aattagctca tcaagaaaga aaaatttata tgggttattc ccagtagaaa 1620
ttaaacagct tactaaatcc tcgcttacat taactgtgta gcttttccct ttattttcac 1680
tgactattgg atagtattca ggataataag aacaataaca aactcatatt gtgcctggct 1740
cttttctaaa tactttacat atgttaccta atttagtcct aacaacttag gagataggtt 1800
gttattaatg gtgcttgtat agtactagca tcatcagtag tagtagtgat agtagtagtt 1860
attactactt cattacaact tttagttatt acaatattat aatgttgttc tcatcatttc 1920
tagataggta aactaaggca ttaaagttta agtaacttgc ctctaaaact atacagctcc 1980
ctgatggctt acaaagacat aaaataagat atacttacca aatgttaagt taaataccta 2040
ttggcaaaag taatgctttt acagccagtt agattattta acagcttgtc acatatatac 2100
accaaggaca tcatcaacct gtcttttcaa aattgtaaga gaaagaccct tgaattcctg 2160
cagtgctagg taatgcaatt aagtgtttgc taaactatcg ggcataagag cgacttcttc 2220
tatctctggg ttgtagcaaa acatataact gctcagatag gatataaatg agctgtaatt 2280
tcctaactgg ctttttacat ttaccaattc caaatcagaa gtaatgtctc ttcactgggt 2340
aactaaagtg ttccctttgt ctgaactgtt cattcaactc aattagactc ctgaaatcaa 2400
ttgttggctt tcacctatgt gtttatcttc atagactttt catatttggg tggtaatctg 2460
gacaggaaac tttagcaagt cacacatgga tgagaaaatg ttgaattaaa taataacttt 2520
caaaggaacc aataatttat tgagtactta ctatatggta ggcactgtgc taagtggttt 2580
attaaccctc ttttatgaat acagaaatta aagcaaagag cagctaagta actttgtcca 2640
aggtcacata gctagttagt ggcagagtta gaattctatt cctttaaaat agctatgtct 2700
aatattattc aattgttttc agttgtgtga acttttttagt aaactagtcc agaatttttat 2760
caggtggagt gctttagatg taagcttatc taatgacatt gatacaaatt acagattttc 2820
tggaagaacc tcaaatatca tctggtccag gtttttgttt tattttaagc tgtgttccac 2880
agatctctag aagtttcgtg gaagatactg ggaggggaaa taggggttga gaaagactaa 2940
aagtgctaat gagtaatttt taaaaggcat tactacaaga gattaagcat tctcctgtca 3000
caattaagaa tttatactac gatatctatg tgttctgtgt agtcaataaa aacattgtct 3060
tttagctctg aatgatttga gcaaggtttc tatccaataa ctaagaacaa agatttcata 3120

acacacattt tattttctct aagtgtaggg atgaaataat cttaatgatt tgttgtttgt 3180
tgttaaatgg aatgtttgca ttctgtacca aagactctaa aattaagttt tagtatattt 3240
gtacataaaa ttatggaatt taacatttgg gccaaaattc tgaatgtaat acttttgtca 3300
aaaacttttt ttaatgtgtg ggggaaagaa ggaagagatg atactctact ctgagtgttc 3360
agaccatttc aaagtatctt atagctatta taaatactta taaagactga ttaatataaa 3420
aattcaacaa aactattaaa tgagagaagg cagtgtttaa gagtatggtg tctggaggat 3480
atagtcctgg tcctgaattt actgagtgag aagaggatgt atgtcatcaa ctcttgatta 3540
gccgactgta cttgagcaag tcagcctctc tgagcctcag tttcctcacc tgtaaaacaa 3600
gtgtaataac agagcctacc tcatagcatc atcctatttg taaggattaa ataaaacaag 3660
tgtataaagc acagtagttg gcaatgtagt aaacacttta taaatgttaa ctattgttgc 3720
cattattatt tttcatgttt aaaaacttag atcacaaaca caaagaaaaa aattgttttg 3780
gtgaatggct gcatcctgtc tttgccagct gaagataatt aagagatcag taattcatca 3840
atcaggctag cgaatttata tcctaaaatt gtatgtgatg gcactttaaa tcagcataac 3900
ataacagaaa aaaaaaccct tcagttttcc tgtaaaactt tactgcattt cccccacacc 3960
tcagtgtttt gattttcctt ttgccaaagg cgatccaccc ttcctgctgt atctattatc 4020
agactccatt cttcttcctg cctcccaccc ttaatcatgt ttccactcac taaacctagt 4080
tttgattgga tccttagtct gacttctatt acaaaacaaa tgcagctggt aaggttgggt 4140
tgcctcttcc ccattcctct tcacaccctg ccatcataaa gatcaacaat atcattttct 4200
ttgtcacatc cactatcagg gaaagaaaat tttgtcaaaa aattgaaatt ttgtccagtg 4260
tttctggacc ttaataattc tacgcataat agatcagagc agccgtaaga tgaagtacct 4320
tttatttcct tctataggct actctctcta gtctttccta tcataattct tggtgatttt 4380
aatatctaca cagatgattc ttccaacact ctagcccctc agatccctga ctttccctcc 4440
tccagggatc ttagtcctca tcatctctca ggtgcttctt cccatagtca tacgcttacc 4500
tttgtcattg caatgtctgc aacctctgca taatatcatt tatttggggg tgtttttgt 4560
tcttcttttt gaacttctct attttcatag gtacatgttt aactttgaca aaatacttta 4620
aaaagcagtt gtaccatttt acacttcact tcattatgtg agagttccac ttgctccact 4680
ttcctgtcaa cacttggtat ggtcattctt tttcatttca gttattctaa tgtgtttatc 4740
atggtatctc attgtggttt taatttgcct ttcccacatg tctaatgata ttgggcatct 4800
tttcatgtcc ttatttatca tctgtatatc ttcctttgta aagttttcaa atctcttccc 4860
cattttaatt gttctttaac tttaattttt aatttttgtg agtacatagt aggtatatat 4920
atttatgggt tgcatggaat attttgatac aggcatgcaa catgtaataa tcacatcagg 4980
taaatgggat attcatcccc tcaagcattt atcttttggg ttacaaacaa ttcaattata 5040
ctgttttagt tatttttaaa tgtacaatta aattattttt cactgcagtc accttattgt 5100
gctagcaaat actaggtgtt attcatcctt cctagctatt ttttgtaccc attaacactc 5160
ttcacctccc cacacacaca gactcactac ccttcccagc ctctagtagc catcctttac 5220
tctctctatg agttcaaatg tttttgttct tagctctcac aaataagtga gaacacgtga 5280
agtttgactt ctgtgcctgg cttattttat gtaatatatg acgtccagtt ccatccatgt 5340
tgttgcaaat gactgaatct cattcttttt tatggttgaa tagtactctg ctgtgtatat 5400
gcccacattt tctgtatcca ttcatctgtt gatgggatat ttaggttgct tccaaatctt 5460
ggctattgtg aatagtactg caatgaatgt gggagtgcnn nnnnnnnnnn nnnnnnnnnn 5520
nnnnnnngct gagatgatat ctcattgtag ttttgatttg aatttctctg atgatcaatg 5580
acattgagca ccttttcata tggctcttca ccatttgtat atcttctttt gaggaatgtc 5640
tattcacatc ttttgcccat ttgtcaaaca cagtattaga ttttttccta tagagttatt 5700
tgagctcctt atatattctg gttattaatc ccttgtcaga taggtggttt gcaaatactt 5760
tctcccattc tgtgggtcgt ctttgcacat tgttgattcc tttgctgtgc agaagcttgt 5820
taacttgatg tgatcccatt cgtccatttt tgctttgctt gcctgtattt atggcatatt 5880
attcaagaaa tctctgccca ctccaatgtc ttggagagtt tccctaatgt tttcttttag 5940
tagtttcata gtttcaggtc ttagatttaa gcctttaatc catttgtatt tgatttttg 6000
tatatggtga gagatagggg tctagtttca ttcttttgca tatggatatc cagttttccc 6060
agcacctttc cccagtgtat gatcttggca cctttcctga aaatgagttc attgtaaatg 6120
tatagactta tctccaggtt ctctattctt ttccactgat ctatgtgtct ttttttatgc 6180
caggaccatg ccattttggt tactatagct ctgtagtata atttgaagtc aggtattgct 6240
taggagatag ctttggctat tctgggtctt tcctggttcc atataaatat taggattttt 6300
aaaaatttct gtgaatatat gtctttgtca ttttgatagg gattgcatta aatctataga 6360
ttgctttggg tactatggac attttaacta tatttattct tccaattcat aaatatagaa 6420
tatctttcca tttttgtatg tctttttcaa tttcttgtat caatgtttta tagttttcag 6480
gtagaaatct tttagtattt ttgttaatta ctatgtactt tatttcattt gtagctattg 6540

caaatggaat tactttcttg attttttttc acattgttca ctgttggcat atagaaatgt 6600
cactgatttt tgtacgttga ttttgtaacc tgcaacttta ctgaatttat caactttaag 6660
agttttcatt ggagtgttta ggtttttcca aatataagat catatcatct gcaaacaagg 6720
taatttgact tcctcctttc caatttggaa gccttttttat ttctttatct tgtctgattg 6780
ctctggctag gacttccagt actatgttga ataactgtgg tgaaagtggg catccttgtt 6840
atgttcccaa tcttagagga caggatttca gtttttgtcc attcagtata atactagcta 6900
tgggtttgtc atatatggct tttattctgt tgaggtatgt tccctctata cccatgtttt 6960
tgagggtttt ttgtcataaa gggatgttta atattatcaa atgctttttc agcaacaatt 7020
aaaatgatca tgaggttttt gttcttcatt ctgttgatat gatgtatctc attaattgat 7080
gtgtgtatgt tgaatcattc ttgcatcact ggaataaatt gcacttggtc atgataaatg 7140
atcttttgtt ttgtttttgt tttcactttt aagtacaggg gtacatgtgc agatttgtta 7200
tataggtaaa cttgtgtcat gggtgtttgt tgtacaaatt atttcatcac ccaggtatta 7260
agcctagtac ccattagcta ttttttttcc tgagtccatg tattctcatc ttttagctgc 7320
cacttgtaag tgagaatgtg tggtatttgg ttttctgttg ctgcattaat ttgctaggga 7380
taatggcttc tagctctgtt catgttccta taaaggacat gatctcattc ttttttaaaa 7440
aagtgacttt attttatttt agttacataa attacaaaat atcactaagt gaaaataaaa 7500
tcaataaaaa tcatccatga taccacccac tttaacattt atgtgtatag cctcttatgc 7560
tttatttcct cacatatata gataaataca ttcatcaaaa agaggttatt tcatatatta 7620
agttggtaca aaattaattg cgcttttttgc cattactttt aatacattgt tttgcaaact 7680
atttttattt cacaatatat tatgaattta tttctataac attaaatata ttatctgtat 7740
gtatgtgtgt cttggtttct tatgactaaa atttttttaaa attaaggcgt tgttatgttg 7800
agatagttga agattcacat gcagttttaa gaaataatac agagagatgc tgtgtgccct 7860
ttaccaagtt tcccttaatg ataacatctt gtaaaactat agtatgataa gaaaaccagg 7920
acattattga cattgatgca gtcagataca gaagattttc atcactacac agatccgtgt 7980
tgccctttta aagccacatc cacttgtgtc tcatccagtc cctcaaccat taatctcttt 8040
tctgctttga taattttatc atgtcaagaa tgttacgtaa atggaataaa acagtatatc 8100
acctttttggg attgtctttt ttttccccac tcagcacaat tccctggaac ttcatccaag 8160
ttgttgtgtg tatcaacagt ttgatccttt ttattgctga gtactactcc atgatactga 8220
tatgccacag tttgtttaat tattcagctg ttgaaggaca ttttggttgt cactagtttt 8280
gggttattac aaacaaggct gtataaatcc tcttttacag gtttctttat ggacataagt 8340
tttcatttct ctgagataaa tgccaaagga gtctagttgt ttggtcgcgt agcagctgca 8400
tgtttagatt tggaagaaat tgccaaagtg ctttccagtg tggtcatact attttacatt 8460
aaaaccagca ctatttctgt gcattcttac cagcattttg tgttgtcact attattatct 8520
taactatttt gaaagctgtg tagtgacatt ttattgttta aatttgcatt tcccaaaagg 8580
ctaataaaat tgaacatttt gtctgcttat ttgtcatctg catgtcctct tcagtgcaat 8640
gtctgtccat gtcctttgct cattttcttt tttccttttt tattttagag tatttagttg 8700
gcaaataaag attgtatata ttcaatgtat acaacacaat gattttgttt ctttttaaaa 8760
agaattattt atttttcaat gggtttttgg ggaacaggtg aagtttggtt acatgaataa 8820
gatatatagt agtgatttca gagatattgg tgcatccgtc acccaagcag tgtacactgt 8880
acccaatgtg tagtctttta tcttcactcc cccacccctt tctctgagtc ttcaaagtcc 8940
attgtatcat tcttatgcct ttgcgtcctc atagcttagc tcccaattat aagtgagaac 9000
ataccatgtt tggctttcca ttcctgagtt acttcactta gaataatagt ctccaattcc 9060
ttccaggttg ctgcaaatgc cattatttag tgccttttta tggctgagta gtattccatg 9120
gtgtgtgtat gtgtgcatat atatatatat atatatatat atatatatat atatatatat 9180
atcacatttt ctttttttatt atactttcag ttcttggatg catgtgcaga acgtgcaggt 9240
ttgttacata ggtatacatg tgccatggtg gtttgctgca cccatcaacc cgtcatctag 9300
gttttgggct ccacatgcat taggtaatgc tctccctctc atttcccccc actccgcaac 9360
tggctccagt gttccatgtg ttctcattgt tcagctccca cttatgagtg agaacatgtg 9420
gtgtttgctt ttctgttcct gtgttagttt gctgagaatg atgggttcca tctttatcca 9480
tgtctctgca aaggacatga actcattctc ttttatggct gcataatatt ccatggtgta 9540
tctgtgccac attttctttg tccactctat cattgatggc cctttgggtt gattccaagt 9600
ctttgctctt gtaaatagtg ctgcagtaaa catatgtgtg catgtgcctt tatggtagaa 9660
tgatttataa tcctttgggt atatacccag taatgggatt gctgggtcaa atggtatttc 9720
tgattctaga tccttgagga attttcacac tgtcttccac aatgattgaa ctaattgaca 9780
ttcctactaa cagtgtaaaa gtgttcctat ttctccacat cctctccagc atctgttgtt 9840
tcctgacttt ttaatgattg ccattttaac tgatgtgaga tggtatctca ttgcggtttt 9900
gatttgcatt tctctaatga tcagtgatga ttagctttta ttcatatgat tgttggccac 9960

ataaatgtct tcttttgaga cgctcatatc cttcacccan nnnnnnnnnn nnnnnnnnnn 10020
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 10080
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 10140
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 10200
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 10260
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 10320
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 10380
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 10440
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 10500
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 10560
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 10620
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 10680
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 10740
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 10800
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 10860
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 10920
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 10980
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 11040
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 11100
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnngctatgg 11160
gtttttcatg gatagctctt attattttca gatacattcc atcaataact agttgattga 11220
gagtttttag catgaaggta tattgaattt tatcgaaggc tttttatgca tctattgaga 11280
caataatgtg gtttttgtca ttggttctct ttacatgctg tataatgttt attgatttgc 11340
atatgttgaa ccagccttgt atcccaggga tgaagccaac ttgatcatgg tgaataagct 11400
ttttgatgtg ctgctggatt cagtttgcca gtattttatt gaggattttc acaacaatgt 11460
ttatcaggga tattggcgtg aaatttttctt tttgtgtgtg tgtctctgac aggttctggt 11520
gtcaggatga aattggcatc ataaaatgag ttagggagga gtccctcttt ttctattgtt 11580
tggaatagtt tcagaaggaa tggtaccagc tcctctttgt acctctggta gaatttggct 11640
gtgaatccgt ctgatcctgt gctttttttgg ttggtaaact gttaattgct gcctcaattt 11700
cagaacttgt tattgttcta ttcagggatt gacttcttcc tgtgtccagg aatttatgca 11760
tttcttttag tttttctagt ttatttgtgt agaggtgttt gcagtattat ctgatggtag 11820
tttgtatttc tgtgggatga gtggtgatat ccccttttatc atttttttatt gtgtctattt 11880
gattcctctc agttttcttc tttattagtc tggctagcag tcaatctatt ttgttcatct 11940
tttcaaagaa ccagctcctg gatttattga ttttttttgaa ggatttttcg tgtctctatc 12000
tccttcactt gtgctctgat cttagttgct tcttgtcttc tgctagcttt tgaatttgtt 12060
tgctcttgct tctctagttc ttttaattgt gatgttaggc tgttgatttt aaatctttcc 12120
cgctttctga tgtgggattt tagtgctata aatttccctc taaacactgc tttagttatg 12180
tcccagagat tccggtacat tgtgtctttg ttctcattag tttcaaagaa ctttatttct 12240
gccttaatgt cgttgtttac ccagtagtca ttcaggagca gttgttcagt ttccatgtag 12300
ttgtgcggtt ttgagtgagt ttcttaatcc tgagttctaa tttgattgca ctgtggtctg 12360
agagactgtt tgtcatgatt tccattcttt tgcatttgct gaggagtgtt ttacttccaa 12420
ttatgtggtc aattttagaa taagtgtgat gtggtgctga gaggaatgta tattctgttg 12480
atttggggtg gagagttctg tagatgtcta ttcggtctgc ttggtccgga gctgagttca 12540
agtcctgaat atccttgtta attttctgtc tcgttgatct aatgttgaca gtggggtgtt 12600
aaagtctccc actattattg tgtgggagtc taagtctctt tgtaggtctc taagaacttg 12660
ctttatgaat ctggttgctc ctgtattggg tgcatatata tttaggatag ttagctcttc 12720
ttgttgcatt gattccttga ccattatgta ataccccttct ttgtcttttt tgatctttgt 12780
tggtttaaat tttgctttat caggaactag gattgcaacc cctgcttttt ttttctttcc 12840
atttgcttgg taaatattcc tccatccctt tattttgagc ctatgtgtgt ctttgcacat 12900
gagatgggtc tcctgaatac agcacactga tgggtcttga ctctttatcc aatttgccgg 12960
tctgtgtgtt ttaattggaa catttagccc atttatattc aaggttaata ttgttatgtg 13020
tgaatttgat cctgtcatta tgatgctagt tattttgctc attagttgtt gcagtttcnn 13080
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 13140
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 13200
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 13260
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 13320
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 13380

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 13440
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 13500
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 13560
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 13620
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 13680
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 13740
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 13800
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 13860
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 13920
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 13980
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 14040
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 14100
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 14160
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 14220
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 14280
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 14340
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 14400
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 14460
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 14520
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 14580
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 14640
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 14700
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 14760
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 14820
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 14880
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 14940
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 15000
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 15060
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 15120
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 15180
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 15240
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 15300
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 15360
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 15420
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 15480
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 15540
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 15600
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 15660
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 15720
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 15780
nnnnnnntat aaggtgagag atgaggatcc agtttcattc ttctacatgt ggcttgccaa 15840
ttatcccgac attatttgtt gaatagggtg tcctttcccc actttatgtt tttgtttgct 15900
ttgttgaaga tcagttggct ataagtattt ggctatattt ctgggttctg tattcagttc 15960
cattggtcta tgtgtccatt tttataccag taccatgctg ttttggtgac tatagcctta 16020
tagtctagtt tgaagtcggg taatgagatg cctccagatt tgttcttttt gcttagtctt 16080
tcttttgcta tgtgggcttt tttggtttca tatgaatttt agaattgttt ttctagttcc 16140
gtgaagaatg acagtggtat tttcatggga attgcattga atttgtagat tgcttttggc 16200
agtatggtca ttttcacaat attgattcta cccatccatg agcatgggat gtatttccat 16260
ttgtttgtgt catctatgat ttctttcagc agtgttttgt agttttcttt gtagaggtct 16320
ttcacctcca tggttaggta tattcctgag ttgttcattt tattttattt tttgcaacta 16380
tggtaaaagg gattgagttc ttatttttatt ctcagcttgg tcactattgg tatataggag 16440
agctactgtt ttgtgtacat taattttgta tcctgaaact ttgctgaatt tatttaccag 16500
ttctaggagc tttttggatg agtctttaga gttttctagg tatacaaaca tatcatcagc 16560
aaacaggaac agtttgactt cctctttacc aatttggatg cccttgattt ttttctcttt 16620
tctgattgct ctggctggga cttccagtac tatcttgaat ataagtggta aaagtgagca 16680
tcattgtctt gttccagttc tcaggggaa tgctttcatc ttttccctgt tcagtataac 16740
gttggctatg ggttcgtcat agatggcctt tattacctta aggtatgttt cttctctgcc 16800

aattttgctg aaggttttaa tcataaagag atgctggatt ttgtcgaatg ctttttatgt 16860
atctattgag atgatcatgt gatttttgtt tttaattatt tctatgtggt gtatgacatt 16920
tgttaacttg cagatgttaa accatccctg catccctggt atgaaactca cttgatcatg 16980
gtggattatc tttttgatat gctgttggat ttaattagct agcattttgt taaagatttt 17040
tgcatatatg ttcatcatga atattggtct gtagttttct ttttttatgt ccttccttgg 17100
ttttggtatt agggggatac tggcctcctg gaatgattta gagataattt ccttttttatc 17160
caatggaata gtgtcaatag gattggtacc aattcttctt tgaatgccag atagaatgca 17220
gctgtaaatc tgtctggtcc tggacttttg ttgttgttgt tggcaatttt taaattatca 17280
ttttaatctt gctgcttgtt attggtgtgt tcagagttac tataacttcc tggtttaatc 17340
tagaagatct ttgtatttcc aggaatttat cctcttctct aggctttcta gtttatgcat 17400
gtaaagatgt tcacagaagc cttaaataat ttttttgtat ttctgtcgta tcagtagtaa 17460
tatctctcat ttcatttcta attgagttta tttggatctt ctctcttctt ggttaatctc 17520
actaactgtc tatcaatttt atttatcttt tccaataaca agcttttgtt tcacttatct 17580
tttgtgtctt gtttgtttgt ttcaatttca cttagttctg ctctgatctt tatttctttt 17640
cttctgctgg gtttgggttt ggattgcttt tgtttcttca gttctgtgag gtgtgacctc 17700
agattgtgta tttgtgctct ttcagacttt ttgatgtagg catttaatac tatgagcttt 17760
ccttttagca ccactctgat ggttaatact gagtgtcaac ttgattggat tgaaggatgc 17820
aaagtattga tcctgggtgt gtctgtgagg gtgttcccaa aggagattaa catttgagtc 17880
agtgggctgg gaaaggcaca cccaccctta atctgattgg gcagcatctt attagctgcc 17940
agcatggcta gaatataaag taggcagaaa aatataaaaa gatgagactg acttagcctc 18000
ccagcctaca tctttctctc gtgctggata cttcctaccc tcaaacattg gactccatgt 18060
tcttcagttt tgggacttgg cctggttctc cttgctcctc agcttacaga cagcctattg 18120
tgggaccttg tgatcatgtt agttaatact taataaacta ataggatata tataatatat 18180
atcctgttag ttctgtccct ctagagaacc ctgacaaata cagccacttt tgctgtatcc 18240
cagaggtttt gataagttgt gtcactgtta tcgttcagtt caaacaattt ttgatttcca 18300
tcttcatttc attttgaccc aacaatcatt caggaggtta tttaattttc aggtatttgt 18360
gtggttttga ggattcctta tggagtttat ttctaatttt attccactgt gttctgagag 18420
aatacttgat ataattttga ttttcttaaa tttactgaga cttgttttgt gccttatcat 18480
gtggtctatc ttggagaatg ttccatgtgt tgataaatag aatgtatatt ctgcagttgt 18540
tgggaagaat gttctgtaaa tatctgttaa gtccatttgt tttagggtat agtttaagtt 18600
gatggtttat ttgttgactt tctttttga tgacctgtct agtgctgtca gtagagtctt 18660
aaagtccccc actattattg tgttaccatc tatctcattt cttaggtcta gtagtaattg 18720
ttctataaat tcaggagctc tggtgttagg tgcatatata tttaggattg tgatattttc 18780
ctgttggact agtccttta tcattcttta atgtctctgt ttgtcttttt taactgctat 18840
tgctttaaag tttgttttgt ctgatataag aatagctact tctgctcact tttggtgtcc 18900
atttgcatgg aatatctttt tcctttacct taagtttacc tgagtcctta tgtgttaggt 18960
gagtctcctg aagacagcag aaacttgttt ggtgaattct tatccattgt agaatggatt 19020
ctatatcttt taagtgaggc atttaggcca tttacattca atgttagtac tgagatgtga 19080
ggtactattc tattcatcat gctatttgtt gcctcaatac cttggttttt ttttcattgt 19140
gttattgtta tatagatcct gtgagatttg tgctttaagt aggttccatt ttggtgtatt 19200
tcaaggattt gtttcaaaat ttagagctcc ttttagcagt tcttgtattg ccagcttggt 19260
agtggcgaat tctcccagca tttgtttgtc tgtaaaagac tgtatctttt cttcatttat 19320
gaagcttagt ttcactggat acaaaattct tgggtgataa ttgttttgtt taaggaggct 19380
aaaaatagga ccccaattcc ttttagcttg tagggtttct gctgagaaac ctgctgttac 19440
tctgataggt tttcctttat aggttacctg atgctttttc ctcatagctc ttaagattat 19500
tttgtttgtc ttgattttag ataacctgat gactatgtgc ctaggcaatt atctttttgt 19560
gataaatttt ccaggtgttc tttgagctta ttttatttgg atgcctagat atctaaaggc 19620
tggagaagtt ttccttgatt attccctcaa atatggtttt cagacttta gatttctcct 19680
cttccttggg aacatcaatt agtcttaggt ttgggtgttt aacatagttc caagcttttt 19740
ggaggctttg ttcacttttt ttgtttttta attatttttt tctttgtctt tgagggattc 19800
agttaatttg aaagccttat cttcaagctc tgaagttctt tctcctgctt gtttgattct 19860
actgctgaga ctttccagtg cattttgcat ttctataagt gtgtccttaa tttccagaag 19920
ttgtggttgt tttttattta tgctatctat ttcattgaag atttttgctt tcatatcatt 19980
gaagattttt cctttcatat cctgtatcat gtttatgatt tctttaagtt ggagttcacc 20040
tttctctgat gtctccttga ttagcttaat aatctacctt ctgaattctt tttctggcaa 20100
ttcaggtatt ttatcttggt ttggatccat tgcttgtgag ctggtgtgat cttttgggag 20160
tgttaaatta ccttgttttg ttgtattacc agaattgatt ttctagttat ttctcacttg 20220

cttagactat gtcagaggga agatctggga ggggctgttc aaattctttt ctcccacggg 20280
gtggtccctt gattgttgtt ctcacacttc ctctagaatt cgggcttcct gagagccgaa 20340
ctgcggtgat tgtttttgct cttctaggtc tagccaccta gcggagctac tggcttcagg 20400
ctggtactgg agagtatctg caaagagtcc tgtgatatga cccatcttca tgtcttttgg 20460
ccatggatac cagcacctgc tctggtagag atagcagggg agtgaagtgg attctgtgag 20520
agtctttggt tgtattttta tttaatgtgc tggatttgta ttggttggcc tccagccagg 20580
aggtggtgct ttcaagagca catcagttgt agtagtctag ggaggaagca aactttccct 20640
agggtcacct ggttaagtat tcaggtttct cgggtggtgg gcagagccat agagctccca 20700
agagattatg tcccttgtcc ttgcaaccag ggtgggtaga gaaagaccac caagtggggg 20760
cagggttagg catgtctgag ctcagactct ccttgagtgt agcttgctgt ggctgctgta 20820
ggggatgggg gtgtggttcc caggccaatg gagttatgtt cccacgggga taatagctgc 20880
ctctgctgag tcatacagat caccaaggaa gtaggggaaa gctggcagtc acaggctcat 20940
cccgcaccca tgcagcctgc agtcctaaag gccagtctta ctcccactgt gccccctcaa 21000
cagcaccgat ctatttctgg gaatctggtg atcaaggctg agaacttgcc ccagaccacc 21060
agcctcccag ctaaaaagca aggagactca cagtttttca gcatctcagg gagcatgcag 21120
cagtgctcca gttccttcaa agggtctgtg gattatctca gcttccctgg aatgttgctg 21180
tggtacttct tggagcaaaa gatcatgatg tgagcctcca cacctctctg tctgtccaag 21240
tgggagctgc aagctagtgc tgcctcctat ctgccatctt aattatctta ttctttttta 21300
tggctgcatc atattcagtt gtgtatatgt accacattgt ctttatccag tctaccattg 21360
atgggcattt aggttgattc catgtttttg ctattgtgaa tagtgctgca gtgagcatgt 21420
gtgcatgcat ctttatgata aaataattta tatctctttg ggtagatacc cagtaatagg 21480
attgctgggt aaaatggtag ttctattttt aggtctttaa gaaaatgtca cactgctttc 21540
cacaatagtt gaactaattt agactcccac ttaacagtgt ctgtgttcct ttttccctgc 21600
aactttgaca gtagttttgt ttttttttt ttttttttgc cttatttata tagagaggtg 21660
gcattttgct atgtatcctg ggctggtcta gaactcctgg gctcaagtga tccatcctcc 21720
ctccgtggca tcccaaagtg ctaggattgc aggcatgagc catggtgccc agcctatttt 21780
tgacttttta atcatagcca ttctgactgc gtgagatggt gtctcattct ggttttgatt 21840
tgaatttctc taattatcag gggtgttgaa ctttttttc atacgctcat tggccacatg 21900
catgtcttct tttgaaaagt gtctattcat gttatttgcc cacttttca taggcttttt 21960
cttgtaaatt tgtttaagtt tcttatagat gctggatatt aggccttcat cagatgcata 22020
tggaaaaaca ttccatgttc atggatagga aaaataaata tcattaaaat ggccatactg 22080
cccaaagcaa tctatagatt caatgctatt cctatcaaac taccaatgac attcttcaca 22140
gaaccagaaa agactatttt aaaattcata tggaatcaca aaaagagccc aaatagtcaa 22200
agcaatccta agcaaaaaga acaaagctgg aggaatcacc ttacctcact caaaactata 22260
ctacagagct atggttacca aaacagcatg gtactggcac agaaacagac acatagaaca 22320
atggaacaga atagagagcc caaaaataag gccacacacc tacaacaatc tgatctttga 22380
caagcctgac aaaaacaagc attggggaaa agactcctta ttcaataaat ggtgctggga 22440
taattggcta gccctatgca ggaggtttaa aatggacccc tttcctacac catatacaaa 22500
aataaactca agatgggtga agtactgaaa tgcaaaatgc aaaagtgcaa aaaccctgga 22560
agaaaaccta ggcaatacca ttctggacat aggaacaggc aaagatttca tgatgaagac 22620
accaaaaaca actgcaacaa aaggaaaaat tgacaaatgg ggtctaatta aacttaagag 22680
ctcctacaca gcaaaagaaa ctatcaacac agtaaacaga caacctacag aatgaatgat 22740
cattttaata tgttgttgaa ttcagtttgc tagtatttta ttgacaattt ttgcaacaat 22800
aatcatatgg tttggctgtg tccccaccca aatttcatct tgaattgtag ctcccataat 22860
tccctcatgt tgtgggaggg acccagtggg agataattga atcatgggca cagtttcccc 22920
catactgttc tcatggtagt gaataagtct cacaagatct gatagtttta taaggggaaa 22980
ccgctttccc ttggctctca ttctcttctc ttgtctgctg ccatgtgaga tgtgcctttc 23040
accttctgcc atgattttga ggcctcccca gccacaagga actatgagtc cattaaacct 23100
ctttcttttg taaattgccc agtgtcgggt atgtctttat cagcagcatg aaaatggact 23160
aatacagtaa attggtacca agaatagggt gctacttaaa agatactcaa aaatgtggaa 23220
gcaactttgg aactgggtaa taggcagagg ttggaacaca tgggagggct cagaagaaga 23280
cagaaaaatg tgggaaagct aggaacttcc tagagacttg ttgaatggct ttgaccaaaa 23340
tgctgatgat atggacaata aaatacaggc tgaggtggtc tcagatggag atgaggaact 23400
tgctgggaac cggagcaaag gtgacacttg ttatgtttta gcaaagagac tggcagcatt 23460
tttgtcccct gccctagaga tttgtggaag tttgaacttg agagagatga tttagggtat 23520
ctggcagaag aaatttctaa gcagcaaagc attcaagagg tgacttgggt actgttaaaa 23580
gcattcactt ttaaaaagga aacacagcat aaaatttcag aaaatttgca gcctgacagt 23640

gtgatagaaa agaaaatccc attttctgag gagaaattca agccagctac agaaatttgc 23700
ataagtaaca aggagcagaa tgttaatcac caagacaatg gggaaaatgt ctccagggca 23760
tgtcagagac ttttgtggca gcccccttcca ccacaggccc tgagacctaa gaatgaaaaa 23820
tgattctgtg ggctgggcgc agggtccctc tgctgtttgc agtctaggga cttggtgccc 23880
tgcatcccag ccactccagg catgactaga agcggccaaa gtatagctca ggctgtggct 23940
acagagcatg caagccccaa gctttggcag cttccatgtg gtgttgagcc tgcagtgcac 24000
agaagtcaag aattgaggtt tgggaacctg tgcctagatt tcagagaatg tatggaaata 24060
cctggatgtc caggcagagt ttgcttcggg gtggggccct catggagaac ctctgctggg 24120
gcagtatgga agggaaatgt ggggttggag cccccacaca gagtccccat ggggtgctgc 24180
ctagtgtagc tgtgagaaga gggccaccat cctccagacc ccagaatggt agatccactg 24240
acagtttgca ccatgtgcct ggaaaagcca cagacactca atgccagcct gtgaaaacaa 24300
ccaggagaga ggctgtaccc tgcaaagcca caggggcaga gctgcccaag gaaacaaggt 24360
gagaaaaatg caaatgcaag tgtcaggatg gaccaagtgg ccagggcata gccaatccat 24420
tcagtgatct cactggggaa attggcttca gaaacataca taaacaagcc accttgtgga 24480
ttcctatagg ttatttctcc aggcttcctg acctggcact atatacagtc actataaatg 24540
ttgatttcca ttcccaaaat aaacaagaag acacctaagc taaaccttat aaacccaaga 24600
caatgggaac ccatctctca agcatcagca tgacccagat gcaagacatg aagtctaagg 24660
agatcatttt ggagttttaa gatctgactg ccctgctgga ttccagactt gcatagggcc 24720
tgtatcccct ttgttttggc caatttctcc cattggaatg actgtgttta cccaaatgtc 24780
tgtacccctc cattgtatct aggaaataac taacttgctt ttgattttac tggttcatag 24840
atggaaggga cttgcattgt ctcagatgag actttggatt gtggactttt gagtaaatgc 24900
taaaatgagt taagactttg agggactgtt gggaaggcat aattggtttt gaaatgtgaa 24960
gacatgagat ttgggagggg ccaggggcag aatgatatgg tttggctgtg cccccaccca 25020
aatttcatct tgaattgtaa cacccataat tccctcatgt tgtgggaggg acccagtggg 25080
agataattga atcatgggga cagtttcccc catactgttc tcatggtagt aagtctcatg 25140
agatctgatg gctttataag ggccccttc acttggctct cattctcttc tcttgtctgc 25200
tgccatgtga gatgtgcctt tcaccttctt ccatgattgt gaggccttcc cagccacgtg 25260
gaactgtgag tccattaaac ctctttttat ttttattttt tttgtaaatt gctcagtctc 25320
atgtatgtct ttatcagcag catggaaaca gactaataca aatatttatc agtgatattg 25380
gcctatagtt ttctttttg atgtgtcttt ggttttggta tcatggtaat actggccttg 25440
tagaatgata ttagaagtat tttctccacc tataattttc agaatagttt gagtagaatt 25500
ggtgtgagtt atttttattt ttttatttt gagacagggg ctcactcatg ttgcccaggc 25560
tggagtgcag tggcacaatc ttagctccct tcaaccttga cttcccaagc tcaggtgatc 25620
ctcctacctc agtctcctga gtagctggga ctacaggcac gtgccaccat gcctggataa 25680
ttgtttatat ttttagtaga gacagagggt tttttttact tgtatcttat tgtactgtct 25740
atgtctcaaa acattgttgt agttattatt tttgattggt tcatcattta gtctttctac 25800
ttaagagtag tttacaaacc acagttacag tattataata ttctgtgttt ttctgtgagt 25860
tttatgcctt ctggtgatta cttatttgtc attaacctta ttttttttct gattgaagta 25920
ctcccttag catttcttgt agggtatatc tggtgttgat aaaaagccct cagctttcat 25980
ttgtctggga agatttttat ttctccatgt ttgaaggatg tttttgctgg atatactatt 26040
ctagggtaaa agtgtttttc tttcaacacc ttcactgtgt catgccactc tctcctgacc 26100
tgtaagattg ccactgaaaa gtctgcttcc agacgcactg aagtgccatt gtatgttatt 26160
agtttctttt ctcttgctgc tttaagatcc tttctttatc cttgaccttt gagagttgga 26220
cgttaaatgc cctgagatag tctttttttgg gttaaatcta cttggtgttc tatgacattc 26280
tcgtacttgc atatcaatgt ctttctctag gttttggaag ttctctgttg atatcccttg 26340
aataaacttt ctatcctatc tctttctcta cctcctcttt aaggccaata actcttagat 26400
ttgccctttt gaagctattt tgtagatttc ataggcatgc tttattcttt tttatgattt 26460
ttttcttttt tctcgtctgt gtgttttaaa atagcctgcc ttcaagctca ttaattcttt 26520
cttctgcttg atcaattcta ctattaaaag actttgatgc atttttcggt atgtcagtta 26580
catttttcaa ctccagaatt tccactcgat tcttttaagt tatttcaatc tctttgttag 26640
gtttacctga tagaattctc tgtgttatct caatttttttt ttagtttcct caaaacagtt 26700
attttgaatc tttgtctgaa atgtcacgta tctctgttgc tccaggattg gtccctagtg 26760
ccttatttag ttcatttggt gaggtcatgt tttcctggat ggtcttgatt cttatggatg 26820
tttatctaca tctgggcatt aaagagttag gtatttattg taatcttcac agtctgggcc 26880
tgtttgtacc catcgttctt gggaaggctt taatttggct ttcctgctcc acctctcctc 26940
tcttttgccc aatttatttt aagacagagt ctcactctgt tgcccatgct ggagtagagt 27000
ggcatgatct tggctcactg caacctctgc ctccagggtt caagcaattc tcctgcctca 27060

gcctgccaag tagctgggat tacaggagcc caccaccatg cccagctaat ttttagtaga 27120
gatggggttt catcatgttg ctcaggctgg tctcgaaccc ctgacctcaa gtgatctgcc 27180
tgcctcagcc tcccaaagtg ctaggattac aggcatgagc caccacactt ggcgtctctt 27240
gcccattttt aaagttgggt agttagttgt tgagttgtgt tctttatttg tatttttata 27300
tgttatagat acaggacttt tttatttttct taataattct tttgaaaagc aggacatttt 27360
atttttgctc tatcccagct tattgaattt ttctcttctc tccctcctct gaattccagt 27420
cacattgacc ttctttcagt tctttataca tgccatgctc aagcctattg caagaccttt 27480
gcacatgtta ttccctgttt agaatgccct cttcgtgccc attcatctaa ttaactgtta 27540
cttatccttt gaacttagtt taaatgctac ttcctcaggg aaggccttcc ctgacagacc 27600
ccatatagat ttctcagagt ttctctgtta tacactcata aaatgcactt cctttcttca 27660
aataatttat ctctgtttaa aactgagagt taatttgggg gaatattttt attttaatat 27720
ctggtgtgta tatatatatg tatatgtctg gtatatgtta cacacataat ttgttcagtg 27780
aatattcatt gggtaagtaa atgagtaagt gaagaaagag ggtccaccaa taaactcaag 27840
tgcatataaa atttcaaagc agaaaaagtg ttttccatca gtagaaaaaa tgatggctga 27900
tatagtttag atatttctcc ttgcccaaat ctcatgttaa attttaattc ccaattctgg 27960
aggtggggca tggtggaaag tgtttggatc atgagggcaa acctctcatg gcttagtgct 28020
gccctcatga tagtgagtga gttctcatga gatctggttg ttgtaaagtg tggcatctca 28080
tcccccactc tctctctctc tctctcactc ctgctttcgc catgtgaagt gtctgctccc 28140
agttcaattt ctgctatgag taaaatttcc ctgaggcctc cccagaagct gagcagatgc 28200
tggtgccatg tttgtacagc cagcagaact gtgagccaat taaacctctt ttcttataaa 28260
ttatccagtc tcaagtattt ctttagagta atgcaagagt gacctaatac aatgatgtat 28320
caggctgtgt ttgcaatact ataaaaaaaa tctgagcctg ggtaaattat aaagaaaaaa 28380
agtttaattg actcatagtt ctgcagatat tacaagaagc atggtgctgg catctgattc 28440
tggtgagggc cttaggaagc ttacaatcat ggtggaaagt gaagagggag caggtgtctc 28500
catgctgaaa gtgggaacaa gagagcaagg ggggaggtgc cacatacttt taacaaccag 28560
atctcgaggg aactaactga gcaggaacaa acttattaac aagatgatgg tgctaaacca 28620
ttaatgaggg atccgccccc aggatccaat cacctcctac caggccccac ctccaacatt 28680
ggagattaca tttcaacatg agatttggag gggacaaata tccaaaccat atcagatgga 28740
tttatttaat gaaaggcata agactattaa ctatttgtaa aaatttaaaa atactaaaga 28800
agtcctcata cacttcttac accaaaacaa aatccaaata aatgaaacaa atgcaaaaat 28860
taaaccatga tggtactaga agaaaacgtg atagaaaatc cttatggtaa atcaaaatat 28920
aaaaataaag taaggaaata tgtttttttgt aatcttgatg tatataagca gatcagaaaa 28980
gccagaccac gtagagaaaa agcatggtag atctcatgta aatttaaatt ttacataatc 29040
catgttttta aaattacatg taacatatat cacaaagagt taatgtcttt aaaatacaaa 29100
taatttttcc aaacaataag aaaaagtcat taccttcata aaaaaattaa aaactgtcat 29160
aaacaaacaa ttcacaaaat aaggaaatgg ccaatggcca tatggaaagg caccttcag 29220
aaaggaaatt tggtggaggg aggagccaag atggccgaat aggaacagct ccggtctaca 29280
gctcccagga tgagcgacgc agaagacggg tgatttctgc atttccatct gaggtaccgg 29340
gttcatctca ctagggagtg ccagacagtg ggcgcaggtc agtgggtgcg tgcaccgtgc 29400
gcgagccgaa gcagggcgag gcattgcctc acttgggaag agcaaggggt cagggagttc 29460
cctttctgag tcaaagaaag gggtgacaga tggcacctgg aaaatcggat cactcccacc 29520
caaatactgc gctnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 29580
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 29640
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 29700
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 29760
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 29820
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 29880
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 29940
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 30000
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 30060
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 30120
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 30180
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 30240
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 30300
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 30360
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 30420
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 30480

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 30540
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 30600
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 30660
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 30720
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 30780
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 30840
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 30900
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 30960
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 31020
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 31080
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 31140
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 31200
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 31260
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 31320
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 31380
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 31440
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 31500
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 31560
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 31620
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 31680
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 31740
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 31800
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 31860
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 31920
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 31980
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 32040
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 32100
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 32160
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 32220
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 32280
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 32340
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 32400
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 32460
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 32520
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 32580
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 32640
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 32700
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 32760
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 32820
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 32880
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 32940
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 33000
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 33060
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 33120
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 33180
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 33240
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 33300
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 33360
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 33420
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 33480
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 33540
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 33600
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 33660
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 33720
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 33780
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 33840
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 33900

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 33960
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 34020
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 34080
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 34140
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 34200
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 34260
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 34320
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 34380
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 34440
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 34500
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 34560
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 34620
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 34680
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 34740
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn tctggggact gttgtggggt ggggggaggg 34800
gggagggata acatcaggag atatacctaa tgctagatga cgagttagtg ggtgcagcgc 34860
accagcatgg cacatgtata catatgtaac taacctgcac aatgtgcaca tgtaccctaa 34920
aacttaaagt ataataataa aaaaaaaaga aaggaaattt ggtaagatct atcaaaatgg 34980
gaaatgtgca tacattttac tgaccatttt cattttaaag attaacctta aagatataat 35040
ctcagaagtg gaagaagcta tatgcccaga aatgtttgtt tctgaagtgc ttagagtagt 35100
aataattttg gaatatctta aatgtctatc aataggaaaa ttatataaat tctgataata 35160
tataaaattt attattatta ttatgtaccc atcacagttg taactttaca tataatgaga 35220
ttatttgctt ccctatatct ctctgtccat agatgatgga gtacatgaga ttaagaatgt 35280
ccatgtttgt ctctagcaac tggctcattt cctattagga actaaataca tacttattga 35340
acaaacaaat gaactgaggt actctctttt cttaataggc tggaagtgat ggagagagta 35400
ttggaaactg tccctttgc caacgccttt tcatgatcct ctggcttaaa ggagttaaat 35460
ttaatgtgac aactgttgac atgaccaggt aagagaaatc aggacatgtt aaattctagg 35520
aattgagatt ggtagatacc aataaaatat tggtgtttat ttaatgtgta ctttatctag 35580
agacctaact ctgcttattt ttaataatca tagaaagcct gaagaactaa aggacttagc 35640
cccaggtacc aatcctccgt tcctggtgta taacaaggag ttgaaaacag acttcattaa 35700
aattgaggag tttttagaac aaaccctggc tcctccaagg tacagcattt acaagatact 35760
attttgctga agataatcta ttttactggc ttgtttattg cagatttagt attcttacca 35820
atttaagtac ttttggattt ctgggcctac atgtcaaatg acacacatgc ataaacatac 35880
ccctccaact tcaaatacaa aaagatgata tgtgtaatat ttcaaataat ttttaaaagc 35940
tgcataacat acataacaca agaaggtaag ttctctgtgc tctagaaata gagtaggaac 36000
atatagtgag atgggagtga gggaatggga tactaacact atgtaattca taaggattgg 36060
tcatgactgg tccttaacac cactgacgaa atgacagaac atacccaaca cgagggctag 36120
tggccaggac atagactcaa gcagttaacc agaggccaga actgtactgc cacactgaat 36180
gacaaccgca catctctgtc tacccaggat aggtctagaa acaagaagca tgctgtatta 36240
attttctatt gctgtgtaac aaattaccac aaacttagta tcttaaaaca acagctattt 36300
attatctcac agcttccatt ggtcagttgt ctgggcatag cctgctaagg tcctctgctg 36360
agggtatcaa aaagtggcat tcaaggtgtt ggtcgggaac acagttatca tatggggccc 36420
aggtgactct tccatcttca ttcaagtttt tggcagaatt cagttccttg cagctatatg 36480
actgaggtct taggttattg ggtagctgtt tgttggggtt gggttggcat tcaattacta 36540
gaggctgccc ctctgtatag gcatttcgca acatggctga ttgttctctt cctctaaaac 36600
ctgcaggaga atgtctctct gatggttcac cttctttttt ttttttttt ttttttttt 36660
gagaccggag tctcgctctg tcccccaggc tggagggcag tggcacatgt tggctcactg 36720
caagctcccc ctttcgggtc tcgggttcac gccattctcc tgcctcaacc tcccgagtag 36780
ctgggactac agatgcccgc caccacgccc gggtaatttt ttttttttt ttttggaatt 36840
ttaataaaaa tgagggttca cccggttaac caggatgggc tcaatctcct gaccttgtga 36900
tccacccgcc tcggcctccc aaagggctgg gattacaggc gtgagccact gcgcccggcc 36960
tgatgggtcc ctttctttta aatttttta tcagcacaaa ttatgggata cctatgaaat 37020
tctattatgt gtttgtaatg catagtgata nagtcaaggt atctacggtg tccataaccc 37080
aaatacaata catttttgta actatagtca ccctgctctt ctatcaaaca ttgaatttat 37140
tccttctatc ttatttatgt gtgtactttt taacacactt ctcttcatct tcccttctcc 37200
tcccaatcac cctccccagt ctctgttatc tctctttcca ttctctatct tcatgtgatc 37260
aacttttta actcccacat ataagtgaga acatgctatt tttgtctttt tgtgcctggc 37320

ttatttcact tgacataaca actccagttc catccatgtt gttccaaatg acaggatttc 37380
attctctttt atggctgaat actatttcat tgtgtatgta taccacactt tctttatcca 37440
tttatctgtt gatggacact tagatcgatt ccataccttg tctattgtga ataatgcaat 37500
aataaacatg agagtgcagg tatccctttg acatactgat ttctcgtgct ttggataaat 37560
gccaattagt gagatttttg gatcttatgg tagtgctact tttggttttt tcagaaattc 37620
tccatgcgtt ttccatagtg gctatattta tactgnnnnn nnnnnnnnnn nnnnnnnnnn 37680
nnnnnnnnnn nnnnnnnnnn nnnnnatttt ttttgctgtg cagaagcttt ttagtttact 37740
tgagtcctat ttgtctattt ttgtttctat tgcctgtgct tttgacatct caatcataaa 37800
ttatttgtct agaacaatgt ccagaagaat tttccctagg ttttctctta ttatttttat 37860
agttttgagt attatgttta agtcttcagt ccttttgagt tgatttttgt atacagtgag 37920
agataaggat caagtttcat tcttctgcat atggctgtcc aattttccca gtaccattaa 37980
ttgaaaaagg tgtcctttcc ccaatgttct tgtgaacttt gtcaaagatc agctggcagt 38040
aaatatgtga atttatttct aggttctcta ttctgaccat tgctctgtgt gtctattttt 38100
ataccataac atgctatttt ggttactata gccttgtaat atatttcaaa gtcaggtaat 38160
gtgatgcctc tagctttgtt ctttttgctc agaattgctt tggctatatg gaatcttttt 38220
tggttacatg tgaattttag tattcttttt ttgtaattct gtgaaaaatg acattggtat 38280
tttgacaggg attgcattga atctgtaggt tactttggga aaatcacaat tttaataata 38340
ttcattcttc tgatccatga gcatgagatg ttttcccata tatttttatc attttcaatt 38400
cctttcatta gcattttgta gttttcattg taaagatctt ccacctcctt gattaaaatt 38460
attcctagat attttaattt ttagctattg taaatggaat tgccatcttc atttcttttg 38520
tgggtagatc attattggtg tatagaaatg ctacatattt tttagtgttg atgttttaa 38580
cctggaactt tactgaattt acttatcaaa tctaagaatt ttttggtgga gtttttaggt 38640
tttactagat acaagatcat ggcaccagta aaaagggaca attttacttc ctttttccca 38700
atttggatgc cttttatttc tttctcttgc ctgattgcca tacctaggac ttccaatact 38760
atgttgaata ggagtggtga aagtgggcat tcttgttttt ttccatttct tggaggaaag 38820
gctttcaatt tttccctatt cagcatgata tcagctgtgg gtttgtcata tatagccttt 38880
attattttga catattttcc ttctatgccc catttgttga gaggttttat catgaagggg 38940
tgttgaattt tatcaaatgc tttttctgta tctattgaga tgatgatatg ttttttgtcc 39000
tttattctat ggatgtcata tattgaggtt attgatttgc acatgttgaa ccattcttgt 39060
atcactggta taaatcccac ttgatcatgg tgtattatct ttctgatatg ctattggatt 39120
cagtttgcta gtattttgtc aagagttttt gtatctatgt tcatcagaaa tattggcctg 39180
tagttttctt ctatgtgtgt gttcttgtct ggtttttgta tcagggtggt gctggcctca 39240
tagaatgagt taaggagagt tctctcctct tccatttttt agaatagttt caggagaaat 39300
tggtattagt tcttctggta gaatttgtca gtgaatttgt ccagtcctgt gcttttcttc 39360
attgggagac ttttttatta ctgactcaat cttgctactc attattggtc tgttcatgtt 39420
ttctatttct tcccaattca gtctcagcac attgtatgtt tcctggaact tatccatttc 39480
ctctaggttt atcagtttgt cagcatacag ttgtacataa tggtctctgg taatcttttg 39540
tatttcttac atatatgact taatgtgtcc tttttcattt ctaatttgtt tgtttgggtc 39600
ttctactttt ttggttagtc tagctggcag tttatcaatt taacaaaaac caacttttc 39660
aatcatgatg ctttgtattt tttagtctgt atttcattta gttctgttct ttattacttc 39720
ctttttctgc taatttggta tttggtttgt tcttgctttt ctagcagctt cacatacatt 39780
attagattgt taatttgtca ttttcctact tttttcatgt aggcatttat tgctataagc 39840
ttgcctctta gtgctgcttt tgctgtatcc cacaggttta tgtatgttat gtttcaattt 39900
tcatttgttt caagaatttt ttttcttctt aaattcttta ttgaccattg gttgttcagg 39960
agcatgttgg ttaattttta tgtatttatg cagtttctaa agttcctctt ggtgtttatt 40020
tatagttgat ttgatttcat tgtggcctga gaatatcctt ggtatgattt tcattgtgtt 40080
aaatttattg agacattttg tggcctgaca tatggtccat cctggagaat attccatgtg 40140
ctgatgaatg tatattctgt agttgttgga tagaatgttc tgtaaatgtc tgtttggttc 40200
atttggtcta aagtccagtt taagtctaat gtttatttgt tgattttctg tctagattat 40260
ctatctaatg ttgacagtgg gatgttaaag ttccttccta ttattgcact gcagtctgtc 40320
tctaccttta gatctagtaa tgtttgcttt atgaatctgg atgctccagt attgggtgca 40380
tatatattta ggattgttat atctttttg ctgggttgat ctgtcattat ataatgatag 40440
ttttagtcct tttttcactt tttttgattt aatgtctgtt ttgtcttata tgattatagc 40500
taatcctgct cacttttggt ttccgtttgt gtgaaatatc tttatcaacc catttcagtc 40560
tatatgtgtc tttactagtg aggtgagtct cttgtaagta ctatgtagtt ggattatgtt 40620
ttttaagtct attcatccag tgtatgtctt ttaagtggaa tatttaatct gtttatgttt 40680
acatgtgaag acttatttct gtcattttgt tatttttttc tggttgtttt gtatattctt 40740

tgtttttttt ctctctctct tgtcatttat cattacagtt tggtggtttt gtgtagtggt 40800
aatatttgag tcctttattt tctttatatc caggcaagaa ggatggccac tattctcaca 40860
ctgggagcag tgtataagtg attcagcctt tcctttcttg ttggactcct tacccttcag 40920
acaaattcca catatagcat ttggaatgac tttgggcttg tacccaggaa ctgagttgga 40980
cagtacagaa cgtttgggag catcttttct tggagtggca gctttgtctt ttgatttctg 41040
tcttccctgg aaagagtcct cctggctgtc aggagagctt tccccttcag ataccttgcc 41100
agaggagctg tccgaagtgc ctttatttt ccgcttctca tcacctcaac catcttcgcc 41160
gtcatctgaa tcaccgtcac tgtctagagc agggagctca ggatccagag tggcctcgta 41220
cagggctgtg tttaatggca gataccacag ctcatctggt gagtacctct tataatattc 41280
ctggaactgt ccggggatga gagccactgg gtaagaactg acctttgttc gccctgttgg 41340
caaaacttcc tacttccctt gaggcacctg gataacgtgt ctgcaagtca aaataagctc 41400
ttctttcttc catgcgttcc cggtttaagt tgctgctttt tttggcagct tttttatttg 41460
tttgtttgtt tgtttgtttg tttgtttgtt tgttttgaga cggagtctcg ctctgtcgcc 41520
caggctggag tgcagtggcg cgatcttggc tcactgcaag ctctgcctcc cgggttcacg 41580
ccattctcct gcctcagcct cccgagtagc tgggactaca ggcgcccgcc accacgcccg 41640
gctaatttttt tgtatttta gtagagacgg ggtttcaccg tgttagccag gatggtctcg 41700
atctcctgac ctcgtgatcc gcccgtctcg gcctcccaaa gtgctgggat tacaggcagc 41760
tttcttaata tactcaggca ctttactggc ttcaactttc tgagtattct gttgttacat 41820
ttgggaatac tctttgtaat ggtctgtaat tcgttgacgt tctttttctt gtagaataac 41880
agaatactca gcatgtttgg ctggatattc ctttatcgtt aaaacaatca cttcatcact 41940
gcccagttaa acctagagtg cattgagttt cagtaatgac atttggctct ctcaggtaga 42000
gtttctcctt gtgagacaaa tctcgtctct ctaaatctgg atatttcctt ttaaaggagg 42060
tcacacccaa atattaactg acttgttctt gaagcatata gtattctcct gtttcatcag 42120
gtggccattt gtactctatc aaatttttatg ctggatagta actaaagcca agatcttgac 42180
ttgaagtttc acagctccta gaaccatctc ctgagcccat tcgcctcttt ttggatgcct 42240
gggtccggtc atttgaatta tcttcaattt catccttcga ggactgcgct ccgggggtgg 42300
ctgggtcgct gtcgcatggc cgcggggctg cgggcgggaa tggagaaggt cctgcggcgg 42360
cggcagcgct cctgacatcc gtccgacccc attttttttaa atatctgttt tattcagcat 42420
actctttccc aacatatctg taatactagg cattagcact ctttttacat ctgtgccaat 42480
ataacaggtg tagtgacatc tcattgatac ttaacttgta tttccttgaa tgatagatac 42540
atttaagcat ctttacctat gttgtttgat catttggttt tgttctcctg tgaatcatac 42600
ttgtcaaata ccgtctgatt ttctattagc ttcatactga ataggnnnnn nnnnnnnnnn 42660
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnntttc tattagcttc atactgaata 42720
ggcaaaagct ggaagcattc cactttaaaa gaggcacagg acaaggatgc ccttcttacc 42780
actcttattt aaaatagtat tggaagttct agccagagca gtgaggcaag agaaagaaat 42840
aatagggcat ctaaatagga acatagaaag tcaaactatc cctgtttgca gacgacatga 42900
ttccatatct agaaaacccc atactcggct gcactcagca tcggagccag gagctagtgg 42960
ccgccgccac gtcccaccag acctgcatcc aagcaagtga agatgttaaa gagatcttgc 43020
cagagccaga aatggaaagt acagacctct gaaaatatct attgaaaatg ggcaacttat 43080
gattggatca tatatagtca gccttcagat tcctgggata acgattatga ttcctttgtt 43140
ttacccctgt tggaggacaa acaactgtgc tatatattat tcaggttaga ttctcagaat 43200
gcccagggat atgaatggat attcattgca tggtttccag atcattctca tgtccgtcaa 43260
aaaaggttat atgcagcaac aagagcaact ctggaaaagg aatctggagg tggccacgtt 43320
aaagatgaag tatttggaac agtaaaggaa gatgtatcat tacatggata taaaaaatgt 43380
ttgctctcac aatcttcccc tgccccactg actgcagctg aggaagaatt atgacattaa 43440
atcaantgag gtacagactg acgtgggtgt ggacgataag catcaaacac tacaaggagt 43500
agcatttcct atttctcgag aagcttttca ggctttggaa aaaataaata acagctgaac 43560
tatgtgcagt tggaaataaa cataaaaaat gaaattataa ttttggccaa cacaacaaat 43620
acagaactaa aagatttgcc aaagaggatt cccaaggatt cagctcgttc catttctttc 43680
tgtataaaca ttcccatgaa ggagactatt tagagtccat agtttttatc tattcaatgc 43740
ccagttacac atgcagtata agagaacgga tgctgtattc tagctgcaag agccctctgc 43800
tagaaattgt agaaagacaa ctatggatgt tgtaatggat gtaattagaa agattgagat 43860
agacaatgag gattagttga cttcagactt cctttgtgaa gaagaagtac atcccaagca 43920
gcatgcagga aaaagaagaa ttcgaagact aattaggggc ccagcggaaa atgaagctac 43980
tactgattca agtcatcaca ttaaacatag caatactagt tttttaaaag tccagctttc 44040
aatacaggag aactgaaatc attccatgtt gatataaagt agggaaaaaa ttgtactttt 44100
tggaaaatag cacttgtcac ttctatgtac tttttaaatt aatgttacat aagagtcatg 44160

```
atttctattt ttgacttaaa gctagaaaag agttcaacat aatgtttaat tttgtcacac 44220
tgtttttata gtgttgattc tacactttca catacttgtt aaaattttat acaattgagc 44280
cagttctaga aagtctgatg tctcgaagga taaacttact actttcttgt aggacagaaa 44340
gaccttaaaa tattcttatc acttaatgaa tatgttaaag accaggctag agtattttct 44400
aagctggaaa cttagtgtgc ctcggaaaag gccagaagtt gcttattctg agtagctgtg 44460
ctaactctgt cagactatag gatcatctct gcaactttta gaaatagtgc tttatattgc 44520
agcagtcttt tatatttgac tttttttaa acagcattaa aattgcagat cagctcactc 44580
tgaaacttta agggtaccag atattttcta tactgcagga tttctgatga cattgaaaga 44640
ctttaaacag ccttagtaaa ttatctaagg ctctgtgaag ccaaacattt atgttcagat 44700
tgaaatttaa attaatatca ttcaaaagga aataaaaaat gttgaaagag ttttaaaaat 44760
caggattgac ttttttctcc aaaaccatac atttataggc aaattgtgtt ctttgtcact 44820
tctgaacaaa tattcagatt taaaattact ttaaagtcct agtatttaac aggctaacac 44880
agataaacac cttaataatc tcctttcaat taatattgta tttcaaacca catttaactg 44940
tcttctaatg ctttgcattt tcagttacaa cctagagaga ttttgagcct catatttctt 45000
tgatacttga aatagaggaa gctagaatac ttcatgttta gtctgttaaa cctgctacaa 45060
aaaccataac tttgaggcat tttctaaatg agctgtgggg atccaggatt tgtaatttat 45120
tgatctaaac tttatgctgc gtaaatcagt tatcagaaat gcacatttca tagggtgaaa 45180
cactcatttt ttttttttt gagacggagt tttgctcttg ttgcccaggc tggagagcaa 45240
tcgcacgatc tccgctcact gcaacctctg cctccagggt tcaagtgatt ctcatgcctc 45300
agcctcccaa gtagctggta ttacaggcat gtgccaccat gcctggctaa ttttgtattt 45360
ttagtagaga cggggtttct ccatgttggt caggctggtt gtgaactccc gacctcaggt 45420
gacccgcccg ccttgccctc ccaaagtgct gggattacag gtgtgagcca ctgcgcccgg 45480
ccaaagcact tatttctaaa ccttattatc taaggtaata tatgtacctt tcagaaattt 45540
gtgttcaagt aagtaaagca tattagaata attatgggtt gacagatttt ttatatagaa 45600
tttagagtat ttgtgtgggg ttttgtttgt ttacaaataa tcagactata gtatttaaac 45660
atgcaaaata attgacaata atgttgcact tgtttattaa agatataagt tgttccatgg 45720
gagcacacat ggacagacat acatacaccc aaactattgc attaagaatc ctggagctgt 45780
gttgcagacc atagctgaag cagttatttt cagtcaggaa gactacctgt catgaaggta 45840
taaaataatt tagaagtgaa tgtttttctg taccatctat gtgcaattat actctaaatt 45900
ccactacact acattaaagt aaatggacat tccagaatat agatgtgatt atagtcttaa 45960
actaattatt attaaaccta tgattgctga aaatcagtga tgcatttgtt atagagcata 46020
actcatcatt tacagtatgt tttaggtggc attatcatac ctagacaatg aataacatat 46080
tcccaataaa tttatatagc agtgaagaat tacatgcctt ctggtggaca ttttataagt 46140
gcattttata tcacaataaa aaatttttct caaagaaaac cccatactct caacccaata 46200
ggtccttcag ctgataaaca actttggcaa agtttcagga tgcaaaatca atgtacaaaa 46260
atcacttgca tttctataca tcaacatcag ccaagctgag agcccaattg ggaaggcaat 46320
cccattcaca attgccacac acaaaaaaat aaaatacctg ggaatacagc taactcagga 46380
ggtgaaggat atctacaatg agaattacaa aacactgctc aaagaaataa gagaagacac 46440
aaacaaatgg aaaaatatcc catgctcatg gataggaaga atcaatatca acaaaatgac 46500
catactgccc aaagcaatct aaagattcag tgttatttct aacaaactaa caatgacatt 46560
cttcacagaa ctagaaaaaa ctattttaaa attcttatga aaccaaaaaa gagcccgaat 46620
agccaaggca attctaagca aaaataacaa agctggaagt atcgcattag ccaacttcga 46680
actatactgc aaggctacag taagcaaaac acagcatggt actgatacaa cacctgtaat 46740
ccctgcactt ttggaggccg aggcaggtgg atcacctgag gtcagctgtt ccagatcagc 46800
ctggccaaca tggtgaaacc ccatctctac taaaaataca aaagttagcc aggcttggtg 46860
acacacgcct gtaatcccac ctactcagga gaccaaggca ggagaattgc ttgaacctga 46920
gagatggagg ttgcagtgag ccaagatcac gtcattgcac tccagcctgg gcaacagagt 46980
gaaactctgt ctcaaaagaa aaaaagaaag aaagaaagaa aaaaacaggc acatagacca 47040
atgggacata atagagagcc cagtaataag gccgcacacc tacaaccatg tgatttttga 47100
caaagctgac aaaagcaatg gggaaagcac tccctgttca ataaatggtg ctgggctggc 47160
tagccctatg cagacgattg aagctggacc cgttccttat accatataca aaaatcaaga 47220
tggattaaag acttaagtgt aaaacccaaa actacaaaaa cccagaagac aacctaggca 47280
atgccatcct agacatagga acaggcaaag atttcatgac aaagatgtca aaagcaattg 47340
caacaaaagc aaaaattgac aaatgggatt taattaaatg aaagagcttc tacacagcaa 47400
aagaaacaat caacagagta aacagacaac ctacagaatg gaagaaaatt tttacaaact 47460
atgcatctaa caaaggtcta atatccagtg tctataagga gcttaaataa atttacaaga 47520
aaaaaatcgc attcaaatgt gggcaaagga catgaacaga tgaacagaca tacatggggc 47580
```

```
aaattagcat atgaaaaaag ctcattagtg atcattggag aaatgcaaat caaaaccaca 47640
atgatatacc atctcacaca agtcagaatg gctaaaaata aaaataaaaa gtcaagaaat 47700
agcagatgct ggcaaggttg tggagaaaag caaacactta tacactgtca gtgggagtgt 47760
aaactagtgc aaccattgtg gaagatagtg tagtgattct tcaaagagct aacagcagaa 47820
ctaccatttg acccagcaat cccattactg gatatatacc cagaggaata taaatcattc 47880
taccataaag acacgtgcat gagaatgttc attgcagcac tattcacaat gacaaagaca 47940
tggaatcaac ccaaatgccc atcaatgaca gactgaataa agaaaaggtg gtacatatat 48000
accatggaat agtatgtagc catagaaaag aatgagatcg tgtcttttgc aggaacatgg 48060
atggagctac aggctattat tcttagcaaa ctaacacagg aacagaaatc caatactaca 48120
tgttcgcata tataagcggg agctaaatga tgagaactca tgaacacaaa gaagggaaca 48180
atacacactg gggtgttctt gagggtggag ggttggagga gggaaaggag cagaaaagat 48240
aacaactggg tactgagctt aataccttgg tgatgaaata atctgtacag caaattccca 48300
tgacatgagt tcacctatgt aacaaacctt cacatgtatc cgaaactaaa ataaattttt 48360
ttaatgaaat aaatatggtt tttgggggggc ctcctctttc ggctttggag ccccccctccc 48420
tctgtctcgg tatgggggag tttcttcctt ctgtcttctc ccttccttct tgcctattaa 48480
actctccgct ccttaaaacc aaaataaaaa aaaaagaaag aaagaaatat ggtttttatt 48540
tttctcacat aagaaactca gaatgaacct aggatgatag ctccgtaatt tcattaggga 48600
tttcaactcc taatctttct tctctgccat ccttcaagtg aggcttccag tctcaaagtt 48660
aactcatggt gacaatatgt ctgctggaac tccaggcaac agatctaata tacaagccag 48720
ctctaaggag ttttcacaga agccacaccc aaaaatttcc atttacagct cattgtccag 48780
aggtaattca tgtggttaga tctaagtagt ggtatataag tgtgttatct gccatagttt 48840
gcccctctga ccacccaaat aaatgtatgt atccctcttc tcacatatgg aacacacagt 48900
tactacagtc ggcttaaagt ccagtacctt tggatgatgt gcaatatctc cattagatac 48960
taatggtcag gcagtcaaat atattaaaaa ttatctccac ccactctttg acacacccat 49020
ttttaaaagt gaagattcga taacacccca acaacccact ggttcatact agttcataat 49080
agttaccatg acttgaaaaa ggactgaaat attgtttcta cgttttattg ttacaaacac 49140
tgctaaaagg aattgtcttt ttacaaggcc ctccacaacg gttagtcttc catattgctg 49200
gatatgggaa cccttccata tgaactttgt tttatctact ttttaaaagc cttgtaaaca 49260
ccccacatta atggaaatgg tggagtaggg aattccagaa ctccattctt tcataaaagc 49320
aatgaatagg ctggcaaaac tgtcagaagc aacttttca gaactctgga atctaagcaa 49380
aaattacagc agccaggaga acacttaatg aataaaaaat ttaaatttca gtgagagttc 49440
tgtggcattt ttggttacct tgagaccatc ctccaaccct cagcccatca atagtcttaa 49500
aaatggcagc ttatattgca ggtgcaggtt actggtacca gaggaagcga tattgacctt 49560
attttcaatg aactgtgatt gtgtagtttg acctatctgg tggttccctg aaggattacc 49620
tcaatggttt acctttttat cacctgcact agagcttccc cagggctgag gcaccttccc 49680
tggtgctggt tgtggaaaga attttaaagc aaatgtatta gtcacagcta cacagaacaa 49740
ggaataacat ctgggaaaag caatagacaa atggaaaaat cccaggaagg gccaggcgcg 49800
gtggctcatg cctgtaatcc cagcagtttg ggaggccgag gcgggcaggt cacctgaagt 49860
caggagttcg agaccagcct gaccaacatg gagaaacccc atctctacta aaaacacaaa 49920
attagccagg cgtggtggtg catgcctgta atcccagcta ctcgggaggc tgaggcagga 49980
gaatcgcttg aacctgggag gcagaggttg tggtgagccg agattgcgcc attgcactct 50040
agcctgggca tggacaacaa gagcaaaact ccatctcaaa aaaaaaaaaa aaaaatccca 50100
gggagaaaga ggctgagata cttgggggat gcttagggaa ataatggctt caaaacattt 50160
tatgtattct gaggactata gaagactatg catggaccca tttctagatg tgtgctcaca 50220
aaagaactga gaagactagg ctctcaattc tggctaaatt tcaggcactg cacaagcaga 50280
aaatgaaggc aaaggcagaa ctttaaactg tatagctaag caatgaagga gagccccaac 50340
acagaaccaa ccctcaaaaa ctaagaaagc tttttgtttt catagtttgt ttctttgttt 50400
tgcttccagg agtttaataa aatctctgta aaatcaataa ctgactaaag ctaatggaac 50460
aaatatttca gaggccacac ataccaaaaa aatataggct ttacaaaatt agttaagaaa 50520
attaactaaa ccaacaacaa ccacaataag cagcaacaac aagaccaggg gactgggaga 50580
atcaatcaga tttccagagt ttctacatta taacattcaa aacatctggt tttcaagaaa 50640
aaaaaaaaac tgaggcatgt gaggaaacaa gaaagtatgg caaggacaaa aaaccaaaca 50700
ccgcatgttc tcactcatag gtggaaattg aacaatgaga acacttggac acaggatgga 50760
acatcacaca ccggggcctg tcggagggtg gggagggata gcattaggag atatacctaa 50820
tgggcgcagc acaccaacat ggcacatgta tgcatatgtg acaaacctgc atgttgtgca 50880
catgtaccct agaacttaaa gtataataaa aaaagaaatg aaaaaaatac attgcataga 50940
agaaatacga tcatacattt atagcattta gcacaattcc tgacataata aaatactcaa 51000
```

taaaacaaca acaacaaaaa gaaaaaccca cagctgacat tgtactcaat agtgaaggac 51060
tgaagttttt ccccttaaga tcagaaacaa gacaaggatg ttcattgtgg ttggaaaaaa 51120
taattgatgt aatttcaatc ttcttaagtg gttaagaatt gttttgtggc ctaacatatg 51180
atctatcctg gtgaatattc tgtatgcact tgaaaataat gtgtattctg ctacagttgc 51240
ccaaaatctg gggttgaaga agccagctta gttctgggtc gggcctgaag cctggggctc 51300
tgtgggtcag ccttttttgg actcggttgg agcctggtct gggcctgaag cctgagcttg 51360
aatgggccag cctgaaatct ggggccacca gggatggcct ggagtctgta cccatgaggg 51420
ctgtattgga ggctgaatgt ttggatgctg acctggtacc tgtggccatg ggggccagcc 51480
tggagctgag gtccatgggt gtcaacgtgg cactgggaca gacccaaagc ctgggagtgt 51540
gaaggccagc ctggagctga gttggtctgg atactgggtc tgtgggtatt ggccttaaac 51600
tggggtccaa aggtgctagt cttgtgatgg agagggcctg aaagctgagt ctgggggtac 51660
agtggctgtc ctgaagcaaa ggggctgtct tggaggggtg caagcctgga ggtatgatct 51720
ggtgctgaag gaagtctgga gtctggggct actggcccag ggctgggaga ctacatctgc 51780
agggatggcc tggacattgg ggctacaagg gctggcctac tgcccaagtc tgtggggacc 51840
agcctaaagt ctggggtaat catggcctgt ccagggctag actttactgt gttgggccca 51900
gtgtttgggt ctgaggcaaa gtctggtgtt cacttacctc ttcttctccc aagcaaaggg 51960
catctctctc catactgtgg gttggagaag gcataacaca ggtaatttaa aactgtcctg 52020
ctaaggtgaa aaataaagca aaaaagagaa gtagtgatgt tagggaaagg agtgatgttg 52080
caacgttaca attgagcgtc cagagaaagg cttcacttag aaagagatac ccatgaaaaa 52140
gacctgaaag aaaagtggga gcaagggatg tccatgtgtc cccctcacct acgggcagac 52200
caagttaaaa ggctctgggg taggagcttt ccaggcctat ttgaatggta gcaagaaggt 52260
ctgtgtcata attgagcgag tgagggatat gagagaagag aggtaaggtg ggatcacatc 52320
atgtggatcc ttataggcta ctgtaatgag ttaggctgtg actcggtaag atgagacgac 52380
tgcagactac tgagtagggg aaagccatca ctctggcttc tgggtggtta atagactggg 52440
tgggaaagaa ggtggttcat atcatgtggg tccttgtaga ccactatgag cacttgggct 52500
ctaactctga gatgaggaca ttgcaggcta atgagtaggg gaaagacatg acatgactta 52560
cattttaaca tgattgctct gtctatgggt ggagaatatt ccaggtgtat gagggacaag 52620
tatgggaata gggagaatag tcaggaggct gttacagtaa tataggcttt ggactgggca 52680
ggggcgcggg ggtggacaga ttctggatac attttgaaag gtaagctgac cagagttgct 52740
aatagatcaa atgtggagtt agaaggaaag agaggaatca aggaagatac ctaagttttt 52800
gacctgacca tttctagctt ccagtgaatt tttttttatg aaaaggaatt gagtgtttta 52860
gcctttgttt gtattgtata tatttaaggt atatcacatg atgtcttgat atacatatat 52920
atagtgaaat gattactaca gtcaagtaaa ttaacatatc catcgcttca tatagttatc 52980
ttttttatat ggtaagagca cctaaaatct acccctttgca aattttcagt atacaatatt 53040
attagtcctc atattataca ttatatcttc tagacttact cattctacat aactgcaact 53100
ttgtaccctc gacctacatc tccctctttc ctacccccac tgacccggta atcactgctc 53160
tattcttttt tctatatatt tgacctctta aagatgccac acataagtga gatcatggag 53220
tatttgtctt tctgtgcctg gcttatttca cttaacataa cgtcctccag gctcatccac 53280
gttgttgcaa atgacaggat ttcattcttt ttaaggctga ttaatattct attacatata 53340
tatatatata tatatatatc tcacaatttc tatatccatt catctgttga tgggaactta 53400
ggttgtttct atatgttagc ttttgtgaat aatgctgcag tgaacatggc agcacagata 53460
tctccatgag gtgctgattt tttattgaat acttttctgc atctagtcat tatcaaatgg 53520
gttttcttat ttgatttgtt aatgtggtga attatattgg ctacttttt cccatttct 53580
ccatcctatt tattccacca tttgttttat aagttgtaat atttgaaacc atatttttct 53640
ttttcttttt ctttttttga gactgagttt cacttgtccc ccaggctgga gtgcaatggc 53700
gcaatctcag ctcactgcaa cctccacttc ccagcttcaa gcaattctcc tgcctcagcc 53760
tcccaagtag ctggaactac aggcgcccgc caccacgccc agctaatgtt tgtattttta 53820
gtagagacaa ggtttcacca tgttggccag gctggtctca aactcctgac ctcaggtgat 53880
ccacccacct cagcctccca aagtgctggg attacaggca tgagccactg cgcctggcca 53940
aaaccatatt tttctactac tcatgtctgc aaatgtattg tactgacatt atatcttctg 54000
acaaataggc ttttaggagc aagtatggaa accaccattt gaaacattgt ttctacagat 54060
aaatgagctt tggattccag acaactgatt accctgtgaa ctttagaaac caaagtgttc 54120
tgagattgga aaaaatataa acttctactg agagacttct aagggtgttt agtttccagc 54180
acaatgttcc agaacttcca ttttcagtat agtgcaagct agggcacctg gtctctgtca 54240
tgttatgtgc aaatgatagt tgacgcatgt ttctttttaa ggtaccctca cctgagtccc 54300
aagtacaagg agtcttttga tgtgggctgt aacctctttg ccaagttttc tgcatacatt 54360
aagaatacac aaaaggaggc aaataagagt aagatacctt ttctttaaat ctctatttt 54420

ctctcactct tcatcttctc actcagcaaa aatagaattt tcctgaatat atagtatatt 54480
ttggggactg gcctagtctt cccctcattc tctatactct cctctgaaat tccctcgcat 54540
gaagttgtat tagatttaga actcaagatt caatatagct attaccaacc atagctcaat 54600
tagaatattg acatactagg tgtgaactaa ctgcaggact gtgtaccttt aaggtttctt 54660
aaactgtggc acctaccatt tcccatgaac attcttaaat agatttatta tcctctgagt 54720
cacaagaact gtgtttttc tttcactttc taactcttct gatcactttt ctttctttct 54780
tttactctcc tgccaatgca cctccctaag aaaagcccaa aagattaaca ctcactattt 54840
catcttactt tgtcttatca gtgagtagct gagcattcta aatagttaac tagatattga 54900
agagccagtg taagtagtat gtatagatag aggtgtctaa atgtgtggaa agcatattta 54960
gaatgtattt agtcaaaaga caatacattt acaagtaact ctattacttc attgcctcag 55020
attttgaaaa atctctgctc aaagaattca agcgtctgga tgactactta aacaccccac 55080
ttctggatga aattgatcca gacagtgctg aggaaccccc agtttccaga agactattct 55140
tggatgggga ccagctaaca ctggctgatt gtagcttgtt acccaagctg aacattatta 55200
aagtaagtct ttataaggca ggctgaatgg gtgggagggg tttgccagtt gccagcacaa 55260
agcatagtga ccttccagtg cggtattatt atattatagc tttgtcatta tcatcatcat 55320
catgtgtact atatacatct cttttctctt tagagggaag atccataatg ttctcttctg 55380
ggaagtatta aaacttgttt ctttttttt ctttttgag atagggtctt gctctgtcac 55440
ccaggttgga gtgcagtggc atgatcaagg cttattgcaa cccccacctc tgaggctcaa 55500
gcagtcctcc caccccactc ttgagtagct gggactacag gtgcgtgcca ccacgcctgg 55560
ctaatttttt gtacttttg tagagacagg gtttcaccat gttgcacagg ctggtcttga 55620
actcctgggc tcaagtgatc cgcctgcctt ggcctcccaa agtgttggga ttacaggcgt 55680
gagccaccgt gcccagccaa aacttgtttc tttctttcta aatcagaagg tattttccac 55740
tgtcttattt tgtaataata ttacctattt tacagaattg ttaagagaat taaataaatt 55800
aaagcattta aaatgcttag aacagtgcct agatcataat agggaataac caatttgggc 55860
tattagtatt atgatgaatt aatcataaat ttaataaata tttattgcat agacttacac 55920
agaatttact ctttgagtcc tatgccaaac acagagaata tgtaaagaaa gaagacatag 55980
gactctaaat aaactcttag tctagtcgtg gtggatatgt gctcatttc tgtggttcct 56040
tcctctaaat atagtcataa ttaaatacag aatcaatatc aacatgattg taagcatgta 56100
gttttgtcaa catttgtaga caaaacatca aaatagtcca agattcgtgt ctacttcata 56160
gtttatttta tagtgctttt tgtgtcgata agatgccttt gataatcttg acttctaaga 56220
aacatttcta catagtaggc atattactga tgccttcttt ttcctctttt ttttgcaaaa 56280
ttctaggttg ctgccaagaa atatcgtgac tttgacattc cagcagaatt ctcaggagtc 56340
tggcgttatc tccacaatgc ctatgcccgt gaagaattta cccacacgtg tcctgaagac 56400
aaagaaattg aaaatactta cgcaaatgtg gctaaacaga agagttagga gagctcttac 56460
aggagaaaag gctatatttg tgatcagatt ttacttattg acatattaga aaggtttttg 56520
caaataagaa tatgaaaaat actgtttctt ctatccaact ctcttatgaa aaggaactct 56580
gtattttcta ttagccataa ataatctgtc cactgtattt tacaggtctt catactttta 56640
cttaattttc tttatctgta tggcaaacca ctgcaatcct gaatgacatg gaaagcatca 56700
caatcttttg ccctttgctt gaattcctgg aatgcataca tataagctaa acagatgtct 56760
gcagttataa atgtcataag tagaggtaca atctcaccct gctccttaga aacatttcca 56820
tataaatcgc taaaataatt tcacattttt gttagtttaa tatatacatg agtttatttc 56880
tgatataaat aataaataca gagagtgagc atatcagaga ggcaaattct taaagaatga 56940
tttttaaaat cagctctagg aagagctcaa gatcaattgg tcatagaaca gcatttgacg 57000
cctagaacta tgaccacctc atggtcagag atgagaatgt agcctttgtg accagattat 57060
attatttta aatgaagaag cactcattaa ataaaacata attttaaaaa acaatataag 57120
aaacaaagtc aactgaatct tttattcata gaaatgaaaa ggaaaataaa aactgtggct 57180
gaccaaaagg tcttcttgtt gtccataaaa ggataaggta aacagtcctt agataattac 57240
aaaactttct acaaaagtta aaatgttaca ttactatacg tattcagatt cacttgttaa 57300
agtactctta aatcattcaa atctggaaac aaaagctgaa cttaactctt gctccctcaa 57360
aagagaaaca caagcataag tgcagcttca aaaaaggaaa atattttagg ctttggtgga 57420
agggtggagt ttagataaaa tttaaatgaa gtagcgtttt aataggttca aagaaaagta 57480
aggcaatgag caaactcaaa gtactgtcct tgaaaaccat agagtcaaga taaatgtata 57540
gtgtatggtt aggtggcaga gaaatgcaat catgttgata atctttgaga tacatcctgt 57600
catcagtata tttcagaata catgcaatgc actagcaagt tacaattgat agaatacatt 57660
tgaaatgtta aatgaaataa gccaggcaca gaaagacaaa caccacatga tctcactcat 57720
atgtggaatt ttaaaaagtt gatctcactc atatgtggaa ttttaaaaag ttgatctcac 57780
acaagtagag ggtagaatcg tggttaccag gggctaggga gagaaagaag gcagaggcac 57840

```
tgaaagatgt tggtcaatgg gtataaagtt acacctagga agaataaatt ttggtattca 57900
ccacagtagg gtgactatag caaataataa tgtagcatgt atttcaagat agctagaaaa 57960
gcaggttttt aaatgtcacc acaaagaaat aacaaatgtt tatagtggtg gatatggtaa 58020
ttacgcctat ttgatcatta tactgtgtgt acatgcattg aaacaccaca ttgtatccca 58080
tatatatgta caattatgtg cccattatac atttaaaaaa taaattttaa aaaccttcaa 58140
ttaactcttg gtttaaaaga aaaatataaa ccaaaactac atgatctcta aaacaaataa 58200
tgatgatgta aacacttcat atcagaatcc atgggataaa tataaagcag tgatcagagg 58260
aaatttttata actaaacact gctattagta aaaataaaag attgaaaata aattgattaa 58320
atattgaact aacaaaaatt tttaaaatgt gcacaacaat gtgaatatac ttgacacttc 58380
tcaactctct gcttcaaaat agttaaggtg atgagtttta agctatgtgt ttttaacaca 58440
acttaaaaaa aaatgtccaa atggatcttg gtagagcacc agcaaaaaac agaaagaaac 58500
ttagaataag tacaacaaat taagtaaaag aacacaagag attaacaaaa aaagtaagaa 58560
ttaacaaaaa gaatagaaat agcatagacc tagttaacga atcaaaaccc tttatttttt 58620
aaaagattga taatacagac caaaccatta gctacattaa ttgaaataaa acagagaaag 58680
caaaagtatg caaaataaag aatggggaaa taactattag aagaaattta agacattgta 58740
agagactact ttgcagacct ctgtgcaaac aaatttcaaa atctagatga tagagataat 58800
ttcctagcaa agtaaagatt acgaaaaaca actttattag agatatgaaa attgaagagc 58860
tcaatcttca tagaagaaag agagaacatt ttttaaaaag aagaaataga gaaaattata 58920
aggaactact taccaaaaag tatcaatccc cagatagttt cacagggaaa tgctaccaaa 58980
ctttaaaaga ccatatagtc tcaaagtaac tttgcgaaac agtgtttctt ctggaaaata 59040
taaaacaaaa tataaagaaa ctatacataa atattgtact ctaattggca aagttgtttc 59100
tcaaggggat atgtgtagac aattctgaaa cagccataca tgtatactaa gattgaaaaa 59160
ataagtaaat gaactgtagg tgggaagtac aaataatcaa gaaggctagg atgaactatg 59220
tggtactgga ttcgattcag agacatcggt atgtactcaa gtttaactta atattgatag 59280
aggtgaatag atacaaaaat aattacatgt gcgtatatac atgagtcagt atacatatgt 59340
atagttccta gccctgtgtc ctgagagggc ctagaagcaa tagtacccta gtagcaacaa 59400
gcacacccaa tgctaagacc ttggattcta atatcattct ccaata                 59446
```

<210> 4
<211> 243
<212> PRT
<213> Human

<400> 4

```
Met Ser Gly Leu Arg Pro Gly Thr Gln Val Asp Pro Glu Ile Glu Leu
 1               5                  10                  15
Phe Val Lys Ala Gly Ser Asp Gly Glu Ser Ile Gly Asn Cys Pro Phe
            20                  25                  30
Cys Gln Arg Leu Phe Met Ile Leu Trp Leu Lys Gly Val Lys Phe Asn
        35                  40                  45
Val Thr Thr Val Asp Met Thr Arg Lys Pro Glu Glu Leu Lys Asp Leu
    50                  55                  60
Ala Pro Gly Thr Asn Pro Pro Phe Leu Val Tyr Asn Lys Glu Leu Lys
65                  70                  75                  80
Thr Asp Phe Ile Lys Ile Glu Glu Phe Leu Glu Gln Thr Leu Ala Pro
                85                  90                  95
Pro Arg Tyr Pro His Leu Ser Pro Lys Tyr Lys Glu Cys Phe Asp Val
            100                 105                 110
Gly Cys Asn Leu Phe Ala Lys Phe Ser Ala Tyr Ile Lys Asn Thr Gln
        115                 120                 125
Lys Glu Ala Asn Lys Asn Phe Glu Lys Ser Leu Leu Lys Glu Phe Lys
    130                 135                 140
Arg Leu Asp Asp Tyr Leu Asn Thr Pro Leu Leu Asp Glu Ile Asp Pro
145                 150                 155                 160
Asp Ser Ala Glu Glu Pro Pro Val Ser Arg Arg Leu Phe Leu Asp Gly
                165                 170                 175
Asp Gln Leu Thr Leu Ala Asp Cys Ser Leu Leu Pro Lys Leu Asn Ile
```

```
            180                 185                 190
Ile Lys Val Ala Ala Lys Lys Tyr Arg Asp Phe Asp Ile Pro Ala Glu
        195                 200                 205
Phe Ser Gly Val Trp Arg Tyr Leu His Asn Ala Tyr Ala Arg Glu Glu
    210                 215                 220
Phe Thr His Thr Cys Pro Glu Asp Lys Glu Ile Glu Asn Thr Tyr Ala
225                 230                 235                 240
Asn Val Ala
```